

PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

REVIEW of EDUCATIONAL RESEARCH

Volume V

DECEMBER, 1935

Number 2

EDUCATIONAL TESTS AND THEIR USES

This issue reviews and brings down to date the
Review of Educational Research for February,
1933, entitled *Educational Tests and Their Uses*.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

NATIONAL EDUCATIONAL ASSOCIATION

1201 SIXTH AVENUE, NORTHWEST

WASHINGTON, D. C.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

THIS ASSOCIATION is composed of persons who are engaged in technical research in education, including directors of research in school systems, instructors in educational institutions, and research workers connected with private educational agencies. The Association became a department of the National Education Association in July, 1930.

Officers of the Association for 1935-36

President

PHILIP A. BOYER
Philadelphia Public Schools

Vice-President

HOLLIS L. CASWELL
George Peabody College for Teachers

Secretary-Treasurer

WILLIAM G. CARR
Research Division
National Education Association

Executive Committee

The president, vice-president, and secretary, *ex officio*, and the following past presidents:

T. C. HOLY
Ohio State University

PAUL T. RANKIN
Detroit Public Schools

EDITORIAL BOARD, 1935-36

FRANK N. FREEMAN, *Chairman*, University of Chicago

MARGARET A. NORTON, 464 Riverside Drive, New York City

WALTER E. MORGAN, California State Department of Education, Sacramento
The President and Secretary, *ex-officio*

Active Membership—Persons desiring membership may obtain an application form from the office of the Secretary-Treasurer. This form when properly filled in should be submitted to the Secretary-Treasurer for the consideration of the Executive Committee. Upon approval of the application by the Executive Committee the person applying will be invited to become a member of the Association. The Executive Committee has defined the qualifications for membership as follows:

Membership in the Association is restricted to persons of good ability and sound training who are working in the field of educational research, and who can present satisfactory evidence in the form of published or unpublished studies which show ability to arrange, to organize, and to conduct research investigations and experiments. In addition, evidence of an abiding interest in the field of educational research is essential.

Membership in the National Education Association is a prerequisite to active membership in the American Educational Research Association. Any form of N. E. A. membership—annual, five-dollar, or life—satisfies the preliminary requirement.

Active members of the Association pay dues of \$5.00 per year. Of this amount \$4.00 is for subscription to the *Review*. See back inside cover page of this issue. The *Review* is published in February, April, June, October, and December each year.

Entered as second-class matter April 10, 1931, at the post office at Washington, D. C., under the Act of August 24, 1912.

REVIEW OF EDUCATIONAL RESEARCH

Official Publication of the American Educational Research Association, a department of the National Education Association.

The contents of the REVIEW are listed in the EDUCATION INDEX

Volume V

December, 1935

Number 5

EDUCATIONAL TESTS AND THEIR USES

(Literature reviewed to June 1, 1935)

Prepared by the Committee on Educational Tests and Their Uses: E. F. Lindquist, Paul V. Sangren, Ralph W. Tyler, Charles C. Weidemann, and W. J. Osburn, *Chairman*; with the assistance of F. P. Frutchey, Louise Mahone, and William Maucker.

TABLE OF CONTENTS

Chapter	Page
Introduction	442
I. Educational Tests and Measurements in China, England, France, and Germany	443
A. China	443
CHARLES C. WEIDEMANN, <i>Ohio State University, Columbus, Ohio.</i>	
B. England, France, and Germany	445
W. J. OSBURN, <i>Buffalo State Teachers College, Buffalo, New York, and</i>	
LOUISE MAHONE, <i>University of Washington, Seattle, Washington.</i>	
II. Present Tendencies in the Uses of Educational Measurements . . .	455
PAUL V. SANGREN, <i>Western State Teachers College, Kalamazoo, Michigan.</i>	
III. Objective Achievement Test Construction	469
E. F. LINDQUIST, <i>State University of Iowa, Iowa City, Iowa, and</i>	
WILLIAM MAUCKER.	
IV. Recent Developments in the Written Essay Examination	484
CHARLES C. WEIDEMANN, <i>Ohio State University, Columbus, Ohio.</i>	
V. Achievement Tests in Colleges and Universities	491
RALPH W. TYLER, and F. P. FRUTCHEY, <i>Ohio State University, Columbus, Ohio.</i>	
Bibliography	502
Membership List	521
Index to Volume V	533

Copyright, 1935
By National Education Association
Washington, D. C.

All Rights Reserved

INTRODUCTION

THIS REPORT is organized for the most part in accordance with the plan used in the previous issue of the *Review of Educational Research* for February, 1933, which was devoted to educational tests and their uses. One important change has been made, however. The chapter on basic considerations was replaced by a consideration of tests and measurements in several foreign countries. A precedent for this policy exists in the recent issue of the *Review of Educational Research* devoted to psychological tests. Such reviews are also quite common in foreign periodicals.

Acting upon the suggestions of many of our readers we have increased the amount of critical comment concerning the references quoted.

While there is still some overlapping among the chapters, we think we have succeeded in restricting this feature to a minimum. We trust that we have succeeded at the same time in covering the entire field in an acceptable manner.

The chairman of the Committee wishes to take this opportunity to thank the other members for their faithful cooperation. Thanks are also due to Messrs. Maucker of Iowa, Frutchey of Ohio, and Mrs. Louise Mahone of Washington for their valuable contributions to the success of this number.

W. J. OSBURN, *Chairman,*
Committee on Educational Tests and Their Uses.

CHAPTER I

Educational Tests and Measurements in China, England, France, and Germany¹

THE REPORT that follows is reasonably complete for the countries represented. Occasionally a report is also included from Holland and Poland.

A. CHINA

Tests and Measurements

In China educational experimentation is new. The Chinese have but recently thought of education as an area of specialized and professionalized learning. Furthermore, for a long time the traditional examination system has seemed sufficient without any changes of consequence in its procedure or requirements. Recent contact with other countries, however, has challenged the leadership of Chinese social, economic, cultural, and educational life to such an extent that some significant changes in the procedures of educational tests and measurements have been made. This movement began in 1922 when William A. McCall of Columbia University visited China as director of psychological research under the auspices of the Chinese National Association for the Advancement of Education. As early as 1913, and later in 1921, informal discussions on testing were conducted by such men as Liao Shih Chen, Chen Ho Ching, and Chang Yao Hsian; and in such institutions as the Higher Normal Schools of Peking and Nanking.

By 1923, many tests had been standardized, covering such fields as intelligence, silent reading, auditory comprehension, mixed Wenli and National Language, composition, formal handwriting, running handwriting, fundamentals of arithmetic, arithmetical problems, fundamentals of algebra, algebraic equations, algebraic problems, integers, decimals, general science, Chinese geography, Chinese history, household arts, citizenship, health, and the teaching of English in Chinese schools. Many of these tests were similar to those used in America. To quote McCall, "The tests . . . merit, I believe, the conclusion that in every case they are equal, and in most cases they are superior to the like tests in America."

For a few years both intelligence and achievement testing developed in almost every phase of school work, yielding such results as mental and subject ages, correlations of abilities, objective technics of testing, and comparisons by subjects among individuals, classes, and other groups. In the period just previous to 1925, the emphasis on intelligence, language,

¹ Professor Weidemann reports for China, Mrs. Mahone for France, and Professor Osburn for England and Germany.

and arithmetic tests was most noticeable in the educational periodicals. There were also discussions of principles and technics of testing and measuring.

Between 1922 and 1924, testing in the province of Kiangsi was undertaken. All of the public and private schools gave intelligence and achievement tests. Teachers became aware of pupil differences, special classes for exceptional children were proposed, and a new professional interest among teachers seemed to have been created. From 1925 to 1927, interest in the movement decreased, but from 1928 to 1930 a revival of interest was evident. Since 1930, in such municipal centers as Nanking and Shanghai, the measurement of the intelligence and achievement of students has increased. In some experimental schools established by bureaus of education in cities or districts or affiliated with teacher-training institutions, the outcomes of instruction are regularly tested and measured.

Educational Research Organizations

Many institutions engage in some kind of educational research. In the normal schools, colleges of education, and departments of education in the universities, the teachers carry on research and collaborate with advanced students. In the College of Education of Central National University at Nanking, Professor Hi Wei has published a series of reports on research findings which exemplify educational research in such institutions.

The best known organizations covering the general phases of educational developments in China are: the Chinese National Association for the Advancement of Education, the National Association for Mass Education, the National Vocational Education Association, and the Chinese National Education Association. Each of these bodies has published materials, chiefly of a statistical and survey character, relative to many different phases of Chinese education.

Some educational research has been done in such provinces as Kiangsi and Shantung, and in such municipalities as Nanking and Shanghai. At Pin-Yang and Chow Ping some educational social experiments have been conducted.

Since 1928, the Institute of Educational Research at National Sun Yat Sen University in Canton has been trying to devote its entire program to educational research of a scientific character and to assemble a collection of literature essential to such research. The productions of this institute include studies for the improvement of teaching Chinese, the discovery of better methods and contents for adult education, and the relationship between education and the economic life of the Chinese people. This institute also publishes the *Chinese Journal of Educational Research* and a series of research monographs. Its assembled references include an eight volume *Index to Chinese Educational Periodicals* and 34,000 volumes in Chinese and other languages consisting of books, magazines, reference books, documents, research bulletins, and school textbooks (12).

Sample References to Recent Educational Research

Through the representative of the Chinese Ministry of Education, Mr. Wei Hsueh-Chih, a number of recent documents of an experimental nature have been received (see bibliography). Mr. Chih's statement indicates that through Dr. H. H. Hsiao of the Society of Tests and Measurements at Nanking, an authority on educational research in China, much information may be obtained. The documents received are in Chinese. Many efforts were made to secure their translation, but no person was available who judged himself capable of translating the technical vocabulary. The references indicate something of the nature of the studies, but no evaluation of their content was possible.

Summary

The above description indicates that Chinese educational leaders are developing tests, measurements, and experimentation for Chinese boys and girls. They have established a basis for such development, have succeeded to the point where whole provinces have used educational tests, have started to develop their own institutions adapted to solve Chinese problems, and have begun a movement toward education as a profession. Chinese and American educators should exchange educational research literature in order to further international goodwill and humanitarianism.

B. ENGLAND, FRANCE, AND GERMANY

Educational tests and measurements have developed much further in the United States than they have in Europe. From the European point of view our growth has been too extensive. European thought tends to be organismic and configurational in character. There is little sympathy for our atomism in psychology or pragmatism in philosophy. Educational tests do not thrive where elements or "atoms" are neglected. Pragmatism, with its insistence that anything is good if it works, encourages the growth of educational tests as a means of ascertaining to what extent our desired outcomes have been achieved. In Europe there is little or no emphasis upon educational outcomes as we know them.

Whatever the complete explanation may be, the fact remains that European countries have not shown a marked interest in tests for general use in schools. No large publishing concerns in Europe devote a major portion of their attention to the publication of educational tests and measurements. On the other hand it would be erroneous to state that Europe is not interested in educational tests. Numerous workers in England, France, and Germany are using them for research purposes.

Everywhere during recent years a remarkable interest in intelligence testing has developed. References to intelligence testing were collected through error in gathering the data which relate to France. Rather than lose the results of the work the references are included. All other articles relating *exclusively* to psychological tests were excluded. In many cases,

however, foreign authors make no clear distinction between psychological and educational tests. This will account for several citations that are apparently psychological in nature but which nevertheless devote considerable attention to tests and statistical procedures of an educational type. For this reason a few references are included which occur also in the June, 1935, issue of the *Review of Educational Research* relating to psychological tests.

Criticism of Testing

Since the testing movement is questioned in Europe, one would expect to find many articles of a critical nature. There are a few. Juhasi (87) discussed the effect of the organismic movement on testing. Szondi (185) reported a critical and historical study of psychological tests which is applicable to educational tests as well. Thomson (191) would replace intelligence quotients with standard scores.

Another sort of criticism is directed at the American testing movement as such. Duthil (48) likes what we are doing. Leitzmann (106) also offered favorable comment. He wonders "why we [in Germany] have so long overlooked the achievement tests and the rich literature that goes with them." He then offers these reasons. In general, Germany has paid little attention to investigations in other lands. The great war and the years of inflation following it have interfered. There is an attitude of "What can America, the land of practicality and pragmatism tell to us, the very authors of pedagogy?" The author quotes William Stern to the effect that our tests are too mechanical and that they are the product of a "test-cult." Leitzmann then proceeds to quote several of Counts' adverse criticisms of tests but offers nothing in rebuttal. He sees a contradiction in our claim that every child is unique and test construction based on central tendencies. In spite of all adverse criticism, the author concluded that Germany can gain from educational tests "much valuable help in formulating more objective and useful judgments."

Statistical Procedure

There is much interest in statistical procedures related to testing. Blumenfield (20) gave an interesting mathematical treatment of test evaluation. Bobertag (22) has a rather elementary discussion of statistics for his readers in Germany. Dwelshauvers (57), Meili (115), and Spearman (178, 179) presented discussions concerning the G factor. Emmett (64) and Wilson (202, 203) were concerned with the tetrad criterion for use in scholastic examinations. The point at issue was the G factor. Fessard (71), Levitof (103), and Valentine and Emmett (197) were interested in reliability. Statistical phases of test standardization were discussed by Fessard and Piéron (68), and Mme. H. Piéron (147, 148). Kern and Lindow (89) and Lipmann (108) reported discussions of curves and

curve fitting. The applications of the theory of probability were presented and criticized by Moers (120), Myers (127), Poppelreuter (151), and Thouless (193). Pauli and Wenzl (134) presented a "very simple method for the calculation of coefficients of correlation." They suggested the

formula $R = \frac{M_1 - M_2}{M_1' - M_2'}$ in which the M 's represent the means of the upper and lower halves respectively of each distribution. Rosenblum (161) developed a method of testing homogeneity among test items. Thumb (194) presented clearly the theory and technic of multiple correlation applied to the weighting of test items. He suggested the use of polar coordinates which he illustrated with interesting geometrical figures. Urban (195, 196) wrote a series of articles on methods.

Test Batteries

Ten articles were concerned with test batteries. Ballard (17) is clearly the most influential author of educational tests in the countries covered by this report. His tests are available for commercial use and are commanding attention in both France and Germany. Decroly (37, 38, 39, 40) used them in Belgium and they have been used in Germany. Duthil (51, 55) published a manual for the use of intelligence, achievement, and aptitude tests. Sandon (167) reviewed a secondary-school test prepared by Brockington. Rather elementary discussions of testing methods were presented by Duthil (51, 54, 56), Frickx (77), Nihard (130), K. Stern (182), and W. Stern (183).

Tests as Survey Instruments

A number of articles were concerned with the use of tests for survey purposes. Bartsch (18) compared the achievement of pupils in two types of German schools. His tests were designed to measure concentration of attention, interpretation of familiar pictures, the combination of meaningless figures, the power of imagination with reference to given pictures, the power of observation, and forgetfulness. Bartsch used the profile technic but failed to describe his tests. Bobertag (23) compared estimates and measurements of achievement in the German folk schools. He showed that measurements are much more reliable. Bobertag is one of the few German investigators who has the American point of view concerning tests. A series of articles covered a study of achievement in certain schools of college rank in Germany (83). The authors were interested in the stage of knowledge attained by pupils of different intellectual and social levels. They agreed that "testing procedure is looked upon as an American effort toward rationalization and tends to become the single ideal toward which the human soul strives," but they could not forget the "very critical condition of the 'Hoch-Schule' studies and their immense defects." The authors expected their studies to be criticized as "grab and snatch" pro-

cedures. They defended them not as individually diagnostic but as a "statistical method of searching after knowledge." The experiment involved 1,143 subjects and many assistants. The tests covered such items as power of observation and description, the understanding of textbook content, and the command of a foreign language to be chosen by the subject. The tests of observation and description contained pictures of skulls half human and half chimpanzee or ape. The criterion was a description written by an expert describer of skulls. The test conditions were carefully controlled in all cases.

The authors were much interested in sex differences. When the girls did not do as well as the boys the cause was ascribed to "the greater sensitivity of the woman." In some cases the girls excelled the boys. Concerning this result the authors say: "An explanation of this fact is naturally difficult to give. One might point to the well-known psychological fact that the female sex is inclined to be, for the most part, more frivolous, more eager for praise, more diligent and quite devoted to details and minutiae." When the girls excel, the authors simply cannot understand it or reconcile themselves to it.

In both the English and French examinations the content was a discussion of M. Briand's reply to Mr. Kellogg concerning the outlawry of war. Reference is made to the isolationist scruples of America. A portion of the text reads as follows: "Whatever happens and whatever the other struggles in which France may find herself involved, she can rely, at worst, on American neutrality." Great care was taken throughout the experiment and much attention was devoted to errors. The articles have much to say concerning the difficulties encountered in evaluating answers.

Huth (84) used modified forms of tests prepared by Bobertag and Hylla with different time limits. He was interested in attention, application to details and to relations, association, command of language, discrimination of space relations, concepts of order, ability to make analogies, and ability in mathematical and organized thinking. Attention was also devoted to such items as interest in work, understanding of work, carefulness, and sobriety.

A test battery published by Bobertag consists of six parts, two of which are devoted to reading, two to arithmetic, one to language, and one to spelling (101). The tests may be purchased at one mark per copy from the Zentralinstitut für Erziehung und Unterricht, Berlin W 35 Potsdamer Strasse 120. This is the only instance noted of tests published in Germany for commercial use.

Müller (124) reported a survey of the relation between school performance and social influences. School performance was measured in terms of teachers' estimates. Weigl (199) listed 790 pupils with reference to logical thinking, spatial discrimination, and power of attention. The occasion for the tests was the passage of the pupils from a lower to a higher school.

Tests of School Subjects

Arithmetic—Tests were reported for eight individual school subjects. Arithmetic leads with eight articles. Burt (30) and Duthil (50) reported tests that are available commercially. Korn (91) presented a study of performance and error in calculation. Lindbeck (107) was interested in the work curve of school children. The task was to complete sums in which the digit in the ten's place had been omitted. Rate was measured by having the children work where they were at the end of each time interval. Work curves of individual pupils were shown along with the curve for the entire class. Deviations were interpreted in terms of temperament, intelligence, emotionality, and interest. Experimentation of this sort would be helpful in America.

Révész (157) reported a study of arithmetical achievement and skill in the highest levels of the primary school (sixth grade). The study was carried on in 19 schools of the city of Amsterdam. The tests were a modification of the Ballard tests. They comprised in part 100 abstract examples and 100 problems. The pupils' papers were examined for errors and the types of errors were reported. In problem-solving, hoaxes were introduced such as the following:

1. A boy has to walk 3 kilometers to get to school. He can ride four times as fast on his bicycle as he can walk. How far must he ride on his bicycle to reach school?
2. A man can go from his home to the depot in 20 minutes. His son likewise can make the same journey in 20 minutes. How many minutes will it take them if they go together?
3. If it takes 3 minutes to boil an egg how many minutes will it take if there are ten eggs in the water?

Nearly 60 percent of the pupils failed to see these hoaxes. Only $7\frac{1}{2}$ percent of the pupils could solve correctly time and space problems such as, "My watch gains 4 minutes a day. If I set it right at noon Monday, what time will it be the middle of next week when it is 6 P. M. by the right time?"

Only $16\frac{1}{2}$ percent filled out correctly "The minute hand goes ——— times as fast as the hour hand." Twenty-eight percent solved, "A thirty-five-year-old man is 7 times as old as his son. How many times as old as his son will the father be in 25 years?" The problem "A man walks 6 kilometers per hour from B. Two hours later another man starts after him on a bicycle and rides at the rate of 10 kilometers per hour. How long must the bicyclist ride to overtake the man who is walking?" was solved correctly by 15.6 percent of the children.

Thirty-five percent of the children correctly solved, "A man goes 5 miles north, then 5 miles east, then 5 miles south, then 5 miles west. How far is he then from where he started?" Only 10 percent solved correctly, "Along a 90 meter street there were trees standing from beginning to end on both sides, 6 meters apart. How many trees stood along the street?"

The following are listed as easy examples that first-year pupils ought to solve. The percents following in each case indicate how many pupils really solved them.

1. There are 40 nuts in a saucer. After 5 men have eaten 7 nuts each, how many nuts will remain in the saucer?—50.3 percent.

2. A boy measured a rope with a meter-stick and found that it was 6 meters long. But someone had secretly cut off one centimeter from the meter-stick. How long was the rope actually?—31.5 percent.

3. If $1\frac{1}{2}$ kilograms of cocoa costs 3 RM. (reichs marks) how much does one kilogram cost?—83 percent.

4. I divided five marks between two girls. One got $1\frac{1}{2}$ times as much as the other. How much did the girl get who got most?—53.1 percent.

5. How many apples can I buy for 1.20 RM. at the rate of 4 for 15 pennies? (one RM. equals $2\frac{1}{2}$ pennies)—22.9 percent.

6. A band box and the key for it cost together 2.20 RM. The box cost 2 RM. more than the key. How much did the key cost?—77 percent.

Like many other investigators the author speaks of the errors as being surprising and almost unintelligible.

As in America, many children failed to read all of their problems. This error caused 47.3 percent of the pupils to fail on the problem, "The foot of a hill is 200 meters above sea level. The top of the hill is 400 meters above sea level. How far above sea level is a house that stands half-way up the hill?"

Inexperience with space relations caused trouble with the following problems. The percents show the number of correct solutions.

1. We have 60 centimeters of iron wire. Out of it we make five triangles of equal size and equal sides. How long is each side of such a triangle?—56.7 percent.

2. A rectangle is twice as long as it is broad. Its area is 200 square decimeters. How long is it?—23.4 percent.

3. Around a rectangular flower bed 3 meters long and 2 meters wide there is a grass fringe $\frac{1}{2}$ meter wide. What is the area of the fringe?—10.3 percent.

4. The outside dimensions of a rectangular mirror are 15 decimeters long and 8 decimeters wide. The mirror has a frame $1\frac{1}{2}$ decimeters wide. What is the area of the mirror inside of the frame?—26.5 percent.

The tests also included material designed to test mathematical understanding and insight.

Schmidberger (169) used arithmetical exercises with 1,121 boys and 1,163 girls in and about Jena to test sex differences. Seemann (170) presented an extensive study of the psychology of number and of errors in calculation. This is in all probability the most complete study of errors with whole numbers that has been published in any language. The author classified errors as mechanical, associative, and functional. Mechanical errors were interpreted as usual. Associative errors were due to similarity of sound, similarity of figure, and perseveration. Functional errors were those of operation, logic, and harmful transfer. The author concluded that errors are never accidental. He gave a very detailed presentation of errors in the four fundamental processes.

E. Thomson (194, 189) used arithmetic exercises as a part of a battery of tests to study the efficiency of individual work.

Spelling—Six references relate to the testing of spelling. Brandicourt (27) used tests which require the subject to write the proper names under drawings. Duthil (49, 56) discussed methods of testing spelling. The results of tests in spelling were reported by E. Thomson (189).

Language—Four articles were found which dealt with language testing. Duthil (53) was concerned with the measurement of ability in composition. Williams (201) reviewed the North Hampton Composition Scale. Bobertag has a language test (101).

Reading—Four studies are also presented which involve the testing of reading. K. Stern (182) was interested in beginning reading. E. Thomson (189) reported results with Haggerty's reading tests. Thorner (192) presented a rather extensive study of the psychology of reading. The content of the test was partly sense and partly nonsense material. Reading difficulties were treated in terms of size of type, vocabulary, length of words, and word form. The data were collected by means of the tachistoscope.

Miscellaneous—Bradford (26) presented a study of perspective in geography. Eaglesham (58) used content relating to the geography of Australia in studying retention. Eggink and Bradford (60) presented further arguments concerning the testing of perspective in geography. Fessard and Fessard (72) discussed musical aptitude as shown by the Seashore records. Mainwaring (111) presented his tests of musical ability. Simon (175) presented a discussion of group tests in drawing. Theiss (188) tested the relation of handwriting to the character of the writer. Leopold (102) presented tests of the appreciation of poetry. Brandicourt (27) tested vocabulary. Robson (159) presented a study of vocabulary burden in the first year of French study. She reported that "children with an average age of about 12 and who spend two hours weekly at French are able to deal with a vocabulary of from 450 to 812 words during the first year. Within this range they succeed in correctly recognizing an average of 71% of the words presented but can actually reproduce only 58% of them." The study is both extensive and thorough.

Examinations and Marks

There is considerable interest in the form of examinations. Champneys (32) presented a bibliography of studies relating to the English examination system. Emmett (64), G. Thomson (190), and Wilson (202, 203) were concerned with a tetrad criterion for use in examinations. Farmer (65) presented data on the predictive value of examinations. Kesseling (90) has a study of entrance examinations, especially for teacher-training institutions. He investigated the possibility of substituting newer test forms for the present state examination in Germany. He suggested tests of memory, attention, and the higher intellectual processes. He presented in detail the nature of the tests that he would have used. He recommended that new

tests be used instead of the present examinations. Valentine and Emmett (197) were concerned with the reliability of examinations. Piéron (138) and Sandon (165) presented criticism of examination methods.

Relatively little is being written concerning the use of teachers' marks in testing. Braun (28) and Sandon (166) based their studies on teachers' marks. Sandon (164) also presented a study of marking systems.

Aptitude and Character

Tests of aptitude are not within the range of this number, yet it seems desirable to call the reader's attention to a few studies in that field. Claparède (33) has a book on discovering aptitude among school pupils, and De Montlebert (45) on determining aptitude by tests. Duthil (51) also presented a study of tests of aptitude. Fessard and Fessard (72) were interested in musical aptitude. Fessard (70) discussed the interpretation of numerical results in aptitude examinations. Lahy (97) reported results obtained with the Stenquist Test of Mechanical Aptitude. Mennens (117) has a study of mental aptitude among prisoners.

Some interest has been shown in tests of moral character. Ekenberg (61) presented a review of the character tests published in America, France, and Germany. Henning (82) presented new apparatus and methods for use in testing character. Moers (120, 121) developed tests of moral understanding, and Reynier (158), tests of character.

Physical Tests

There was also interest in tests of movement. Liefmann (104) studied the relation between mental and bodily achievement. Her tests of physical performance include such activities as stringing beads, rope climbing, balancing, rowing, jumping, and tests of suppleness, speed, and muscular strength in arms and legs. Most of these tests were adapted from those of Serebrowskoja (172, 173) of Moscow. Stephenson (181) was interested in tests of motor perseveration.

Diagnosis, Prognosis, and School Readiness

Quite a number of workers claimed diagnostic value for their productions. Brandicourt (27) presented drawings as a basis of diagnosis. Claparède (33) has much to say on the subject. Kovarsky (92, 93, 94, 95, 96) described the use of the psychological profile for diagnostic purposes. Rohrschach (160) discussed the methods and results of psychological diagnosis.

Only three workers made claims as to tests having prognostic values. Bobertag (21) and Lietzmann (105) discussed the comparative prognostic value of tests and school reports. Farmer (65) studied the "predictive value of examinations and psychological tests in skilled trades."

Tests of school readiness are usually of the mental type. Descriptions of a few of them by Bühler and Hetzer (29), Remy (155), Winkler (204), and others (187) are noted here as samples.

Orientation and Social Participation

The use of tests for orientation and guidance has attracted some attention abroad. Ménessier (116) described the use of tests of professional orientation. Similar discussions were presented by the Piérons (139, 144, 147).

Saluschny (163) reported an interesting study of the measurement of social participation among groups of school children. The experiment was carried on in Charkow, Ukraine. The tests consisted of eight tasks. Groups were required to choose a representative to participate in the school conference of the city. They were also asked to choose a director for the arrangement of the schoolroom, to sketch a plan for community work, such as a school picnic or excursion, to select the pupils best suited for each duty, and to pick out the five members of the group who in the opinion of the group exercise the most disorganizing influence. The plan of scoring is explained. The coefficients of reliability ran as high as $.98 \pm .006$. The test was given to 8,000 children.

Psychological Tests

Owing to their organismic views it is natural to expect foreign workers to use psychological tests for educational purposes. Abadi (15) reported a test of personality. Abrahamson (16) and Bartsch (18) presented tests of imagination. Abrahamson (16), Bartsch (18), and Monchamps (122) were concerned with tests of observation. Bartsch (18), Gamsa and Salkind (79), Mme. Piéron (146), and Weigl (199) reported tests of attention. Bartsch (18), Fischler and Ullert (73), Mme. Piéron (147, 149), and Kesselring (90) used tests of memory. Dilger (46) discussed classification by means of the Gaussian curve. Eder (59) undertook the diagnosis of carelessness. He assumed that $LM = SM \text{ times } K$. Where LM is the best possible performance of an individual, SM is the worst possible, and K is carefulness. Letting LN equal the performance under normal conditions and substituting, the author arrives at the formula $LN = SN \text{ times } K$. He considered $\frac{LM}{LN}$ as a carefulness quotient. It is interesting to note that the author's assumptions are related as factors in a product rather than as addends.

Kennedy (88) found that the Downey Will Temperament tests had low reliability. Mira (119) reported a new test for the exploration of affectivity. Mme. Piéron (145) described what she called psychotechnical researches in the school.

Perseveration—Cattell (31) reported p-tests of perseveration. Rangachar (154) tested perseveration among English and Jewish boys. The advantage was with the Jews. Stephenson (181) described tests of perseveration and offered extended discussion of the p-factor.

Fatigue—Foucault (76) wrote on mental work and fatigue. Lindbeck (107) treated work in written calculation in terms of work curves.

Practice—Lämmermann (99) reported a study of the practice effect upon intelligence scores. He obtained practice quotients by dividing end performance by initial performance. His quotients ranged from .63 to 1.26. He found the practice effect so marked as to cast serious doubt upon the diagnostic value of the tests. Kern and Lindow (89) presented a mathematical treatment of practice curves. They used the technic of curve-fitting and derived $y = a \log x + b$, as the general equation for practice. They described the calculation of the constants a and b .

Reasoning—Müller (125, 126) made extensive and careful studies of reasoning among school children. The subjects were aged from six to eighteen. They were required to complete syllogisms by supplying conclusions. The syllogisms contained space, time, larger-smaller, equivalent, whole-part, genus-species, symbolical (algebraic), negative, hypothetical, disjunctive, and causal relations. Individual reactions to the syllogisms were recorded and age norms were derived. These articles are of interest to all who are concerned with the task of teaching children to think.

Weigl (199) used tests of logical thinking. The tests of arithmetical reasoning by Révész (157) have been described previously under arithmetic tests.

Selection of pupils—Considerable interest was shown in the use of tests for purposes of selection. Emery (63) dealt with the social value of selection. Peyraube (136) discussed German methods of school selection. Schlotte (168) reported tests for the selection of the poorly talented. Wellens (200) discussed the selection of the élite.

Summary

The purpose of this chapter has been to supply information concerning the uses of tests and measurements in China, England, France, and Germany. There is plenty of evidence to show that educational tests are in process of development in China. The Chinese movement is the result of American influence.

The tests of England, France, and Germany have also been stimulated in part by American influence, but they also show a point of view distinctly their own. In England, statistics is still a dominant interest. France is interested chiefly in intelligence tests, while Germany has specialized in aptitude testing. In Europe educational tests are used for the most part as aids in research. Many of the foreign studies are more extensive and thorough than those in this country. American workers can derive much benefit by familiarizing themselves with the methods and point of view of their fellow workers in foreign countries.

CHAPTER II

Present Tendencies in the Uses of Educational Measurements

THAT THE TESTING MOVEMENT of this country has experienced phenomenal growth since 1900 is now the common knowledge of all educators. A study of this growth shows a fortunate decline recently in the output of materials for which, in many instances, the motive was to share in the commercial returns of an innovation. The testing movement is now entering upon a period of slower but much more significant and substantial growth. The tests published and the uses to which they are being put are the result of more precise and detailed investigation. Valuable experimentation involving the construction and use of tests is being conducted. Testing, removed from the mysterious category in which it was formerly placed by conservative educators and laymen alike, is coming to be regarded as an integral part of any educational program. There are those, of course, who look with skepticism at some of the applications of tests, but on the whole, tests and testing continue to flourish. The testing movement has made and continues to make lasting contributions to education at all levels.

Perhaps one of the main reasons for this very wholesome attitude toward measurement is that proposed by Lincoln and Workman (232). They say that, at first, testing was largely for the expert, since he alone knew the field, but now tests are devised for use and interpretation by teachers. Many books written in non-technical language are available on educational measurements, and teacher-training institutions deal with them as a regular part of the curriculum. Such textbooks on supervision as those by Burton (210), Carpenter and Rufi (211), Garrison (221), Monroe and Streitz (240); and others include testing as an orthodox procedure in instruction. Other texts such as those by Brueckner and Melby (209), M. Monroe (239), and Gates (222) make all remedial instruction absolutely dependent upon testing. Some of the larger cities maintain bureaus of research whose primary purpose is to deal with problems of measurement.

One may well wonder in what direction we are going and demand proof that tests and testing are valuable and becoming indispensable. An excellent discussion of the "pros" and "cons" of testing in the social studies which may well be applied generally was presented by Kelley and Krey (230). The discussion dealt with such topics as the relationship of testing and instruction, intelligence tests and homogeneous grouping, reliability and validity of measuring instruments, function and limitation of different types of tests, aspects of instruction and achievement measured by objective tests, relationship between objective tests and important aspects of educa-

tion, relationship between individual differences and tests in the social studies, experimental method, and use of tests for appraisal in educational guidance and vocational counselling.

Along with the "slower but surer" progress of testing has come the analytical criticism of specialists in the field who realize that testing is not a cure-all. McConn (235) suggested that one of the reasons why educators have thus far failed to make continued and systematic use of tests is that there is a lack of comparable forms of examinations. The Cooperative Test Service is seeking to meet this criticism. McConn (234, 236) further criticized the much-repeated error "of assuming that results of any single examination or single set of examinations can be regarded as definite, conclusive, and valid as bases for crucial administrative action or crucial administrative advisement." Wood (268:6), in an article on basic considerations in educational testing, stated that "the chief defect in the testing movement has been the neglect of building an adequate philosophy and system of using test results for effective and constructive educational guidance in the largest sense of that word." Limitations and criticisms of mental testing suggested by Horn (227) include such statements as the following: (a) the movement is still in a formative stage; (b) mental ability is still too exclusively ascribed to determined forces; (c) misconceptions exist concerning the degree to which intelligence, as measured by tests, enters into achievement; (d) it is becoming increasingly evident that the prognosis of scholarly achievement may be better accomplished by special aptitude and achievement tests which concentrate on defined areas of subjectmatter; and (e) it is too often forgotten that there are areas of personality and social achievement in which the ingredients are not limited to intelligence. Concerning objective tests, Horn's chief criticism dealt with the placing of a premium on rote learning and a consequent failure to measure the higher thought processes. However, he stated further that "there is no evidence that these defects are peculiarly characteristic of objective tests" and that "the objective examination is demonstrably superior to the traditional examination."

State and National Testing Programs

Serving as conclusive evidence of the growth of the importance of measurement in education are the state and national testing programs which exist at the present time.

A study by the United States Office of Education (255) showed that in 1933 there were state testing programs in twenty-three states of the Union. Segel (253) described one which is especially worthy of note—the academic contest carried on in Iowa and sponsored by the University of Iowa. All schools are invited to participate and a great many of them do. Each school entering on a competitive basis agrees to give the tests to all students taking the subjects listed. The papers are scored by the schools and sent to the contest director who sends back reports which enable each school to

compare its results with those of other schools as well as to compare its own pupils with each other. These test results are also used to determine school marks, to guide individuals, and to ascertain the general level of attainment. Most of the state testing programs are under the direct supervision of the state university or state department of public instruction.

In the bulletin of the Office of Education (255) referred to above and in a study by Chase (215) the following organizations are listed as conducting national testing programs: the Educational Records Bureau, the Cooperative Test Service, the College Entrance Board, Kansas Nation-Wide Every Pupil Contest, and the American Council on Education.

The Educational Records Bureau conducts an annual testing program for independent schools. Its membership has grown from twelve schools in 1927 to over two hundred at the present time. In 1933 it organized a program for public schools under the direction of a Public School Supervisory Committee. Its chief purposes are (a) to score tests and report results and (b) to act as a coordinating center for testing so that results on different tests may be comparable.

The Cooperative Test Service is an organization designed to construct tests. It has been subsidized by the General Education Board for a period of ten years to construct ten or more comparable forms of examinations in the subjectmatter of the junior and senior high-school levels. This Service cooperates with other agencies such as the Educational Records Bureau and state associations by using their tests in national and state high-school testing. For example, the state high-school testing program under the auspices of the Association of Minnesota Colleges has used the tests of the Cooperative Test Service. Tests are constructed in such subjects as the Romance languages, mathematics, science, and social science.

The College Entrance Board was made up in 1930-31 of fifty members, thirty-nine representing universities, colleges, and scientific schools, and eleven representing secondary schools. It carries on a program of testing with the graduating seniors both in the high-school academic subjects and in scholastic aptitude. These results are acceptable to all universities and colleges in the United States. A report (255) of 1932 shows that 19,929 students took one or more of these tests. The examinations are prepared by the Board itself and arrangements for giving and appointments for scoring are made by it. Committees for scoring papers are composed of representatives of many different colleges and universities and secondary schools. All work is done independently of any one institution.

The Kansas State Teachers College Bureau of Educational Measurement supervises a nationwide scholarship test twice each year. The tests are constructed in the Kansas high schools under the direction of the Bureau. Participating schools score their own tests and send tabulations to the Bureau of Measurements which makes a report containing a national norm and the state norms.

The American Council on Education annually sponsors the construction of a new edition of the American Council Psychological Examination. The examination is widely used with high-school seniors and college freshmen. An edition has recently been constructed for use with high-school sophomores.

The Psychological Corporation has conducted a national survey of English usage which has been made a state program in Pennsylvania and Ohio.

"Cooperative" testing as such is not a recent development. Certain states conducted state high-school examinations many years before the standardized testing movement. The gain in favor of this type of testing is due to the distinct advantage it affords over the individual school testing

programs. Douglass (216) listed among favorable advantages that comparable scores are available for a whole area, making possible a better understanding of achievement, and that this comparison provides motivation for pupils and teachers as well as serving as a means of classification. A further step in the direction of providing standards for comparison, reported by Segel (254), is being carried on under the supervision of the United States Office of Education in an attempt to establish equivalent scores among different intelligence tests.

Trends in Test Construction and Use

There are various methods of classification of the uses of tests. For purposes of brief review the following outline will prove helpful:

1. Prognosis
2. Surveys
3. Diagnosis and remedial treatment
4. Instruction, grouping, and marking
5. Experimentation and research
6. Guidance.

Prognosis—Some valuable contributions to measurement are made by those who would predict the success of an individual in a given field. Fritz (220) recently attempted to determine the value of certain tests for predicting college marks and teaching success. He found two tests (Aptitude Test for Elementary High-School Teachers and American Council Psychological Examination) which he considered helpful but by no means dependable in predicting college marks. He also concluded that if the Aptitude Test really measures ability to teach school, those students earning the better marks are better risks as teachers. Segel and Gerberich (252) studied the possible use of the American Council Psychological Examination in differentiating ability to do college work and ability to do work in specific subjects. They concluded that the test should not be used for differential prediction purposes when college marks are the criteria. Finch and Nemzek (219) studied the prediction of college achievement from data obtained at the beginning and end of the secondary-school period. The authors concluded that "it is quite improbable that a combination of marks from a wide variety of schools will afford a satisfactory means of predicting further achievement until marking systems are decidedly improved. The present results do demonstrate that it is possible under favorable conditions to obtain from marks assigned by high-school teachers a predictive measure of distinct value." A worthwhile evaluation of the prognostic value of different types of tests in educational psychology was made by Terry (260). The Van Wagenen Reading Scale in Educational Psychology, Form A, was compared with the Iowa Silent Reading Test and with the Otis Group Intelligence Scale in an effort to determine the relative value of the tests for predicting ability in educational psychology and in teaching. The

criteria of achievement were an objective test, an essay test, and the objective and essay test combined in a single score. The subjects were the members of two classes in educational psychology attending summer school at the University of Alabama. The findings include the generalizations that (a) the Van Wagenen Test is less dependent upon the factor of intelligence than the reading tests considered; (b) there is a considerable degree of validity in the Van Wagenen Scale and it may be a better means of predicting achievement in educational psychology than tests of general reading ability; and (c) the value of the Otis Test as a means of predicting performance should not be overlooked, although a combination of the Otis and Van Wagenen yield a higher correlation with achievement than either test taken alone. A somewhat similar study was made by Torgerson and Aamodt (261) in which the purpose was to determine the validity of two aptitude tests in algebra as compared with an intelligence test. The tests chosen were the Lee Test of Algebraic Ability, the Orleans Algebra Prognosis Test, and the Otis Self-Administering Test of Mental Ability, Higher Examination. The tests were administered to 236 ninth-grade pupils in Muskegon, Michigan. All three tests were found to be about equally valid and effective in predicting grades in algebra. The aptitude tests were about equally efficient in setting up a critical score below which the students' chances for success were slight. The sharpest discrimination was made by the intelligence test, as 22 of the 23 pupils with intelligence quotients below 90 failed in algebra at the end of the year.

Another important type of prognostic test is the reading readiness test. Segel (254) said, "The problem of accurately measuring the readiness of children to read has become more important since the advent of primary curriculums which allow for considerable variation in the time for beginning reading of different pupils. These tests have been found to have scores which have a fairly respectable relationship to later achievement in reading or other first-grade school achievement." A study dealing with informational background of kindergarten children as it affects reading readiness was reported by Troxel (262). Seventy-four kindergarten children were selected according to teachers' classification of background as good or poor. Data were obtained through the health clinic, teachers' or principals' knowledge of the home, questionnaire filled in by parents, and information furnished by the visiting nurse. The summary of the study stated:

The study of the questionnaire brings out facts which seem to confirm the importance of a rich informational background in reading readiness. The rich background children as judged by the kindergarten teachers of Kalamazoo, live in smaller families, have had more opportunities for travel, have richer play experiences, and are surrounded by a reading environment far richer than the children having a meager background. The study seems to justify the conclusion that the technic used in determining richness of background, or a part of it, could be used in mapping out instruction procedure, in grouping children, in enriching the curriculum, and in understanding some of the underlying causes of failure in learning to read.

Surveys—This use of tests is the one with which the state and national testing programs are largely concerned. A survey is conducted primarily for the purpose of comparing schools with schools, cities with cities, states with states, etc. In other words, such a program is concerned with group results only. Douglass (216) discussed the effect of state and national testing on secondary schools. The chief objections relate to the artificial determination of objectives and methods, the fact that all instructional activities are directed toward cramming things measured by examinations, and the overemphasis on traditional subjectmatter. Advantages which Douglass suggested have to do with facilitation of comparison, better basis for school marks, motivation for teachers as well as pupils, training for pupils in test taking, and the stimulation of teachers' interest in objective tests. Lincoln and Workman (232) stressed the fact that results of survey tests are not individual performances but group averages. They did, however, classify the use of tests to single out individuals of the group for further intensive study under the survey heading. Horn (227) emphasized the use of caution in the interpretation of survey tests in the evaluation of instruction by reminding (a) that any test measures only a minor part of the total results of instruction, and (b) that relative performance is dependent on a large number of factors of which the result of immediate instruction is but one. Other references to the survey test were made by Segel (253, 255), Johnston (229), and Allen (205).

Diagnosis and remedial treatment—The use of tests for diagnostic purposes is of decidedly growing importance. Supplementary to this usage is remedial instruction and general improvement of instruction. A diagnostic test in any subject reveals which skills have been sufficiently mastered and which require further instruction. Ramsay (247) reported the results of an investigation dealing with diagnostic testing and remedial teaching in the junior-senior high school. His conclusions were three: (a) teachers need not hesitate to undertake diagnostic testing and remedial instruction even with the limited facilities of the small junior-senior high school; (b) large gains are possible by deficient pupils in the ability to study, in the mastery of fundamentals of reading, and in the acquiring of the fundamental concepts of arithmetic; and (c) remedial instruction, as evaluated by teacher judgment, carries over to the instruction of the regular classroom.

The need for diagnostic testing was discussed by Olander (245) who pointed out the poor ability of teachers to diagnose pupils' errors. He reported an experiment in which arithmetic tests were given to pupils individually. They were required to make their calculations aloud, and these were recorded verbatim. Two papers, representing four skills each, were mimeographed and corrected by forty teachers who had no knowledge of the verbatim record of calculations. Later, these records were made available and the teachers were asked to correct the papers in the light of their knowledge of the pupils' methods of work. It was found that the average percent score was lowered by 6.6. The teachers were asked to diagnose the

errors before they knew the pupils' methods of work. Results showed that errors in long division were diagnosed with a high degree of accuracy, but the average in the diagnosis of the three other fundamental skills—addition, subtraction, and multiplication—was less than one error out of three correctly diagnosed.

Carter (212) reported a study illustrating the diagnostic testing and remedial procedure involved in a case of reading disability. Such topics are discussed as (a) innate ability and factors affecting performance, (b) school status, (c) initial status in reading, (d) diagnosis and suggestions for remedial work, (e) remedial treatment, and (f) results of remedial instruction.

Smeltzer (257) discussed an efficient method of giving and scoring objective tests for diagnosis in the classroom. Seventy-five test items placed on three mimeographed sheets were distributed to pupils. The numbers from one to seventy-five were written on the board by the teacher. When three-fourths of the pupils had finished, the papers for each row, without names, were collected, and redistributed in another row. Incorrect and omitted answers were checked. Then by counting raised hands, the number of students who made mistakes on each question was recorded on the board. Scores were entered on test papers, the papers returned to the owners, and the names written on them.

A study reported by Whitmer (266) evaluated the progress made by college probationers in the years succeeding remedial work. He concluded that there is a probability that, with assistance, the probation student reaches his level of academic performance more quickly than he would without such help.

Practically any good book dealing with tests and diagnosis is replete with case studies and discussions of diagnostic technic.

Instruction, grouping, and marking—Besides motivating individual remedial instruction, testing is used rather widely in homogeneous or ability grouping. Tests of intelligence are used mainly for this purpose. There has been much discussion of the value of this use of tests with none too clear and definite conclusion as to whether or not its advantages outweigh its disadvantages. A study was reported by Sauvain (250) showing the relative attitudes of parents, teachers, and school officials toward this method. He made the observation that the trend seems to be away from grouping. His study revealed that teachers and principals favor grouping more than do parents, and only those parents whose children are in the lower groups regard the system with real disfavor. From the standpoint of the teacher, grouping is a great aid in instruction.

The Tulsa experiment (263) provides that two types of test be given to children at the completion of the second grade, type 1 to divide the rapid from the less rapid learners, and type 2 to test reading, arithmetic, writing, and spelling. The rapid learners are then divided for each subject on the basis of achievement. Achievement testing and regrouping take place at

the end of each semester. The child does not know whether he is in a higher or a lower group; he is promoted regardless of his classification. Ten or more levels in each of the tool subjects are provided and reclassification is easy. The age of entrance to junior high school is the same, under this plan, as under an ordinary graded system.

O'Shea (246) suggested an improved method of administration in the elementary school wherein intelligence group tests are used to determine (a) the mean mental age of the class, (b) who the superior and inferior pupils are *likely* to be, and (c) what the original tentative group should be with which later-formed subject groups may be compared for gaining valuable information concerning individual abilities, interests, and effort. Classifications should include a definite range of intelligence quotient and chronological age.

Woody and others (270) made a study of the effects of measurement on instruction. They concluded that "measurement properly conceived is an integral part of the complete teaching process and as such becomes an important agency in directing the choice of subject matter and methods of teaching."

An aid to instruction is found in the workbook and instructional tests. Van Liew (265) discussed the justification of the workbook and enumerated standards upon which a justification depends. He stated that the workbook can be justified only when the pupil is given something to do beyond merely reading and remembering subjectmatter—only when he is given something to make him understand, appreciate, think, apply, or construct. Some standards upon which the justification of the workbook hinges are: let it (a) guide and aid study, (b) present facts with a view to significant organization, (c) avoid mere syllabus forms and topical outlines, (d) distinguish between working exercises and practice exercises, (e) be constructed with due respect to teaching procedures, (f) parallel a particular text, (g) aid teachers in use of good textbook, and (h) introduce pupils to serviceable ideas of methods and sources making for independent study.

Maxwell (238) presented arguments for and against the use of the workbook as an instructional aid. Arguments for it include the following: (a) the workbook develops initiative and independence on part of students; (b) it presents material in definite sequence; (c) if outlined to accompany a particular text it assures the securing of that point of view and the covering of essential materials as intended by the author; and (d) it reduces the labor of the teacher. Opposing arguments include: (a) the workbook tends to stultify originality because the references are limited to one text; (b) it tends to minimize individual differences; (c) it emphasizes the teaching of the textbook rather than the development of the child; (d) it tends to stress memory work; and (e) teachers follow workbooks too slavishly.

The results of achievement tests are used for marking pupils. This plan yields comparable results from group to group, since standardized tests

provide the norms. Macomber (237) said, "as a measure of academic ability, records of standardized achievement tests over a period of years, together with records showing the actual attainment of educational goals in terms of mastery of units of work, are most meaningful." The advantages and disadvantages of marks, together with the existing needs which can be taken care of by results of standardized tests were reported by Ayer (207). The case against marks shows that marks given by teachers are inaccurate and often based upon items other than the actual trait being measured; overemphasis is placed upon high marks and the disgrace of low marks has been detrimental to pupils; marks put too high a premium on the acquisition of subjectmatter and too little upon the improvement of the child. The case for marks shows that the best progress is made when learners are aware of the rate of their improvement; quantitative marks are essential for classification, educational guidance, and research. The elimination of marks does not do away with failure but merely covers up poor work. Marks should be made more reliable, more specific, and more discriminating and should be used as checks or guides rather than rewards or punishments. The use of standardized tests may well make marks more reliable, specific, and discriminating.

Experimentation and research—Objective tests, because of their superiority to other measuring devices, are a great aid to educational experimentation. With the aid of measurements, the experimenter is able to set up groups which are equivalent in such traits as he wishes to control. The controlled experiment is most widely used in studies where the effect of a single variable is to be measured. Such problems as the value of different methods of instruction or of different textbooks, the effect on achievement of size of class, etc., are typical of the use to which this method may be put. Lincoln and Workman (232) discussed the method, listing several essential points, such as (a) wise choice of subjects, (b) careful selection of place and time of experiment, (d) control of desired experimental factors, and (e) accurate measurement of experimental factors. Several illustrative cases of experimental method were presented.

Robb (248), for example, reported a controlled experiment to determine the results of direct and incidental methods of instruction in the field of character education. The control group was composed of 182 junior high-school pupils and the experimental group of 110 twelfth-graders. Moral knowledge and ethical discrimination were measured, showing relatively small difference in the two groups. The incidental method, i.e., "purposeful instruction in the development of character through the creation of related experiences closely integrated with the traditional divisions of subject matter, and with pupil awareness of such instruction reduced to a minimum," was used with the control group, whereas the direct method, "involving a clearly planned course of study in moral instruction," was employed with the experimental group. All comparisons by means of rating scales were in favor of the experimental group, even though the

difference was small in some cases. "The superiority of the instructed group was greatest on the ethical discrimination test, and least on the ratings given to one another by the students themselves."

Longstreet (233) employed the controlled experiment method in an experiment with the Thurstone Attitude Scales in which he attempted to determine to what degree high-school pupils' attitudes toward patriotism, the United States Constitution, and war are affected by courses in American history and civics. He found that high-school pupils' attitudes are not changed unless the instructor makes special effort to effect such changes. Segel (256) suggested the use of the controlled experiment to evaluate instruction by radio. Such a study, reported by Gordon (224), was carried on at the University of Wisconsin to evaluate the effectiveness of the radio in the subjects of current events and music. Twenty-five rural schools composed the experimental group and an equal number, chosen by the county superintendent, made up the control group. All schools were provided with the same study materials. The control group was taught by the classroom teacher, but in the case of the experimental group, all instruction was given by radio. Results of the experiment showed that the "materials contained in the radio lessons were better taught than without the aid of the radio."

In the one-group method of investigation, on the other hand, a single group is considered. For example, Charles (214) compared the intelligence quotients as shown by three different mental tests applied to a group of incarcerated delinquent boys. He found that about the same difference exists between the result obtained by the Otis and the Kuhlmann-Anderson as exists between that obtained by the Otis and Binet-Simon. There is a much closer agreement between the Binet-Simon and Kuhlmann-Anderson than between either of them and the Otis, which is estimated as yielding results about ten points too high. Similarly, a study was made by Engle (217) of the personality of a group of high-school honor pupils showing that there is no single factor which correlates significantly with scholarship in this group. With the aid of this one-group method, Wrightstone (271) made a study of the correlation among tests of high-school subjects prepared by the Cooperative Test Service. He found that among these tests there is a higher correlation than "can be explained by the hypothesis of a general factor of abstract verbal intelligence."

The correlation formula is often used in this one-group method of experimentation. This procedure has been employed, for example, by Baker and Broom (208) in a study of a criterion for the choice of a primary reading test. Data yielded by results of six standardized reading tests administered to about ninety students were correlated. The tests were found to validate each other very well, as the coefficients of validity ranged from .57 to .84. Also, Stagner (258) made a study of the intercorrelation of six objective personality tests. Correlations were found to be mostly low, with cases of high relationship when identical elements were involved. Moore and Steele

(241) attempted to evaluate six most frequently used tests of personality. These tests were administered to fifty-eight students at Mount Holyoke. General results warrant the inference that "the atomistic method of evaluating personality seems to have very little promise. . . . An entirely different method of approach . . . seems necessary." Jenkins (228) investigated the Standard Graduation Examination for Elementary Schools as a means of predicting success of pupils in certain high-school subjects. This was done by means of correlating test scores on the Standard Graduation Examination and high-school grades. Results showed that the total score predicted success in social studies better than for other fields; "in general, there is a direct proportion between persistence in high school and comparative total score on the test." Seagoe (251) made an evaluation of certain intelligence tests by means of this method. In general, it was concluded that primary tests agree less with upper-grade tests than the latter agree with themselves, and the former will correlate less highly with school achievement than the latter.

An excellent evaluation of the validity of certain questions which purport to measure neurotic tendencies was given by Landis and Katz (231). The Bernreuter Personality Inventory was administered to 184 house patients and 40 out-patients of the Psychiatric Institute of Ohio University at Athens, either in small groups or singly. All of the subjects were "voluntary" patients at the Institute and all were fairly cooperative and interested in the test. The authors concluded that "the personality inventory gives results indicative of poorly adjusted personality when the individual receives a high score; when the individual receives a low score, we are not justified in drawing any conclusion concerning the satisfactoriness of his adjustment to life. On direct validation, we found that about three-fourths of the self-descriptive statements given by the neurotic individuals really agree with the externally ascertained facts concerning those individuals."

On the whole, it may be said that the standardized test has opened an avenue in the educational field which before was hardly accessible because of the subjective methods of research which necessarily were used. This facilitation of research has added greatly to the precision and care with which materials are now devised, and may be said to be responsible for a great part of the changed attitude toward tests and testing. Education is gradually becoming a science and its practices are being based upon carefully determined facts.

Guidance—The topic of guidance is so all-inclusive that it seems wise to consider several different aspects of it. Williamson (267) defined educational and vocational guidance as "a process of assisting pupils to select and become informed about that occupational field which is consonant with their demonstrated aptitudes, interests, and experiences, and to secure training in line with this choice up to the limit of their educability."

In every classroom there are those pupils who do not respond to the normal routine. Their difficulties may be caused by obvious factors or by a

subtle maladjustment whose symptoms are not outwardly manifested. For this reason, guidance is not to be concentrated on vocational and academic factors alone, but is to include all possible aspects of the individual. Fortunately, the progress of guidance programs has been facilitated by the development of tests of personality, attitudes, character, etc., which, although in need of much greater perfection, give promise of real usefulness in the future. The use of such scales, together with other data collected over a long period of time, according to Segel (255) "will furnish better means of predicting success in school or in different curriculums than we have hitherto had." Heck (225) said, "Many attendance problems would never have been problems if a psychological division had been available where a child could have been adequately studied and where an academic program more in line with his capacities, abilities, and interests could have been provided." McConn (235) felt that the introduction of a guidance program should begin in the secondary level because the high schools serve a much larger group. With this arrangement, the college guidance program might be carried on as a continuation of that started at the secondary level with accumulated data available. Evans (218) reported such a guidance experiment carried on in the Pratt, Kansas, High School over a period of three years. The procedure was as follows: (a) the counselor acquired a mass of knowledge about the pupil's interests from observation, interest tests, and personal interviews; (b) the counselor, pupil, and parents met at the pupil's home to plan a schedule for tenth, eleventh, and twelfth grades and this schedule was filed away on a card. The results of such procedure after three years of operation showed: (a) the vocational counselor had become valuable to the school; (b) the pupil chose electives more wisely; (c) parents became a part of the school system; (d) the pupil had incentive for doing better school work because he had a goal to attain; and (e) the principal was given a source of information which enabled him to calculate the demand for certain subjects and plan his class schedules accordingly. Newell (244) discussed the methods of child guidance adapted to a public school program. His purpose was to give a glimpse of the case study method. He described several representative studies considered by the psychological clinic.

In summary, he says, "Our usual procedure with a new case is to undertake intensive work with school, home, and child for a few weeks. It is then often possible to discontinue active treatment, merely keeping the case under observation." In every instance, the accumulation of essential information as a basis of guidance involved the use of tests and measurements.

Controversy exists concerning the validity and reliability of scales purporting to measure traits of character. Typical of the attitude of the opponents is that expressed by Gillingham (223), "To extend the measurement idea to personality is to attempt to harness the infinite." The opponents feel that character traits are too elusive and intangible to offer a reliable result of measurement and that the tests themselves are invalid.

On the other hand, there are those who are more optimistic concerning the feasibility of measuring traits of character. Although they fully realize the shortcomings and imperfections of such scales, they believe that the future gives promise of their usefulness. Stephens (259) asserted that "the very foundation of all sound guidance programs is built on carefully arranged personal data much of which consists of the results of various kinds of examinations or tests." The Seventh Yearbook of the Department of Classroom Teachers of the National Education Association (242) presented a discussion of personality tests and their uses. Especially noteworthy is the bibliography concerning major problems of elementary-school children and secondary-school children. We read here, "An effort to formulate tests and other devices for the discovery and measurement of factors related to personal development and adjustment has gone far enough to give promise of real usefulness in the future." A bulletin of the Research Division of the National Education Association (243) stated that "although progress is being made, it seems unlikely that the status and needs of the whole personality can ever be discovered with the precision and economy of time now possible in the measurement of general intelligence and subjectmatter learning. . . . The foregoing statement does not imply that available testing procedures have little value in uncovering personality difficulties." There follows a discussion of various studies of personality tests and testing. Segel (254) in speaking of the increasing number of attitude scales constructed under the direction of L. L. Thurstone, says, "These scales appear to be valuable instruments in the evaluation of effects of different environments upon people. These scales will make possible a better approach to the measurement of the effect of motion pictures, radio, and newspapers." Other references to personality tests and their uses were made by Tyler (264), Heck (225), Horn (227), Lincoln and Workman (232), and Woody (270).

The clinical approach to guidance is cognizant of the individual as a composite organism, the study of which must necessarily take into consideration many different aspects. The dangers of interpreting results of a single test as final are pointed out by presentday clinicians, psychologists, and psychiatrists. A good description of clinical procedure was presented by Carter (213) who summarized the work of the psycho-educational clinic at Western State Teachers College, Kalamazoo, Michigan, over a two-year period and presented a reading case which he described as being "partially adjusted" at the end of the remedial period. Factors considered by the clinic in diagnosis were family history, developmental and medical history, school history, and clinical data. Faculty members of the Department of Education and Psychology acted as counselors for certain students assigned to the cases, and, in some instances, cases were handled by faculty members alone. Of the sixty children studied and treated at Western State Teachers College during this two-year period, 25 percent were designated as "satisfactorily adjusted," 52 percent as "partially adjusted," and 23 percent as showing no improvement whatsoever. Henry (226) also presented a

description of clinical procedure in a review of a behavior problem considered by this clinic. Robinson (249) discussed the responsibilities of a child guidance clinic in relation to mental hygiene. He views the clinic as a medical institution whose primary responsibility is to the individual child. These clinics should participate fully in the mental hygiene movement, ever recognizing the needs of the community and ever coordinating home and school. Finally, educators will recognize, through clinical experience, that non-promotion is a major mental hygiene hazard, and that it may be necessary to sacrifice grade standards to protect the child's mental health.

Schools for which clinical service is available are indeed fortunate, but the lack of such service does not mean that the clinical approach to a problem cannot be used. For example, we read in the *Seventh Yearbook of the Department of Classroom Teachers of the National Education Association* (242): "The teacher of today must be more than an instructor. He must be to some extent an experimental psychologist, a diagnostician, and a guidance expert. He must be a practicing physician in the realm of personality, interested both in curative and preventive medicine."

An essential part of the guidance program is a cumulative record system for the purpose of presenting college entrance information, counseling during school years and after school years, prediction of personal success, aid in choice of vocation, etc. The American Council on Education (206) has prepared a form for such a system. This form is discussed by Wood (269) who says, "The American Council Committee on Personnel Methods advocates a method identified by its central and unifying instrumentality—the cumulative record of measures and observations recorded in comparable terms in a form convenient for rapid, accurate, and comprehensive interpretation of growth trends, which is to be used carefully and continuously by all who take the responsibility of advising pupils." Significant features of the record pointed out by Wood are: (a) it is arranged in calendar unit columns which facilitates detection and interpretation of trends in development and also gives cross sections of the individual's recorded status at any time; (b) the measures used are comparable from year to year giving meaningful indications of growth in any function tested; and (c) it provides spaces for objective measures of interests and personality traits, for concrete evidence of interests and development, and for teacher's judgment. A form has also been prepared by the Educational Records Bureau shorter than that of the American Council on Education.

Summary

By way of conclusion, it may be said that the summary of the previous pages bearing upon the present tendencies in the uses of educational measurements indicates clearly that tests and measurements for their own sake are rapidly passing from the educational picture. On the other hand, the use of tests as an integral and essential part of educational procedure and research is growing. Using tests for what they may contribute to the realization of the important aims of education and the solution of educational problems appears decidedly to be the modern tendency.

CHAPTER III

Objective Achievement Test Construction

THE LITERATURE on educational achievement test construction during recent years has exhibited a noticeable change in character. Before 1930, writers on test construction were concerned primarily with the comparative merits of the so-called "old" and "new" types of tests, with the statistical technics employed in the gross evaluation of tests, with the preparation, standardization, and statistical evaluation of specific standardized tests, with the development of new test forms, and with the experimental determination of comparative validities and reliabilities of objective tests of various types or forms. Recently there has been some tendency, not as marked as might be desired, to recognize that the statistical technics ordinarily used in test evaluation are far more fallible and less meaningful than had been supposed; that the factors which account for differences between *types* of tests are far less significant than those which account for variations in quality between tests of the same external form; and that much more can be learned about these latter factors through *internal* analyses of test materials—through qualitative and subjective as well as statistical analysis of individual items—than through gross statistical evaluations or comparisons of complete specific tests. Consequently, there has been some diminution in the relative number of articles dealing with studies of comparative validity and reliability, particularly of the old and new types of tests, and of articles reporting the characteristics of specific standardized tests. A much larger proportion of articles than formerly has been concerned with item analysis technics, and relatively more attention has been given to qualitative and logical, as opposed to statistical, considerations in general.

The Concepts of Validity and Reliability

The early literature on testing was characterized particularly by much loose thinking and by many serious misconceptions concerning the nature of test validity and reliability, by undue confidence in the statistical technics employed to measure these characteristics, by serious exaggeration of the relative significance of the reliability characteristic in test analysis, and by what has been described as the "jingle fallacy"—that of naively assuming without adequate evidence that a test really measures what its name implies that it measures. The literature of the past two or three years has contained several discussions which should contribute significantly to a clarification of current thinking in this area. In the opinion of the reviewers, each of the first three or four of the discussions reviewed in the following paragraphs is deserving of careful study by anyone seriously interested in the problems of test construction and use.

In discussing the statistical technics that are currently employed in psychological test work, Thurstone (341) took every opportunity, as he derives and explains the statistical formulas, to draw attention to the limitations and abuses of, and to the misconceptions associated with, the statistical technics considered, and to develop a sound appreciation of the underlying concepts of validity and reliability. The character of his discussion is indicated in part by the following quotations from his preface:

Correlational methods have probably stifled scientific imagination as often as they have been of service. . . . In this country the reliability formulae have become a sort of fetish rather than a tool. . . . The logic of validity and reliability should be regarded as a tool for the investigation of ideas and not as a sort of research pattern which by itself guarantees scientific respectability. . . . The student should be warned that while reliability coefficients are juggled as though they really were available, these coefficients are in themselves more or less in the nature of estimates and make-shifts.

Turney (343) attempted "to justify a single definition of validity and a single criterion for judging validity." He maintained that "validity is by its very nature determinable by no other means (than the judgment of experts) and the only statistical treatment which is essential to the establishment of validity is that which will refine or assist in the consensus of expert opinion." In a critical analysis of existing concepts he drew specific attention to a number of current erroneous notions concerning the nature of test validity. This well-written article should do much to secure more adequate appreciation of the important fact that the judgment evidenced by the test builder in the selection of the elements to be tested and the ingenuity he shows in the construction of the individual items testing for these elements are of far greater importance than the statistical technics he may employ.

Monroe (325) ably criticized the *uncritical* use of objective tests by those who think (a) that complete objectivity in scoring is absolutely essential and that any objectively scored test is an accurate measure of achievement, (b) that, if a test is highly reliable, scores on it are necessarily accurate measures of achievement as specified by the announced function of the test, and (c) that a high correlation with a criterion is sufficient evidence to justify the use of the scores yielded by the test as highly accurate measures of the achievement considered to be defined by the criterion. He pointed out quite clearly the errors in these beliefs and urged more intelligent use of objective tests and more critical interpretation of the scores obtained from them.

Tyler (344, 345, 346, 347) laid particular stress upon the necessity of a careful formulation of the objectives of instruction in a given course before actual construction of a test is begun and upon the importance of building the test so as to measure the degree of attainment of each of those objectives. He used a description of the procedures followed in setting up such objectives in the Department of Zoology at Ohio State University, to illustrate the technic outlined.

Willoughby (353) dealt with the concept of reliability and how to measure it; he stressed the fact that the essential factor in increasing the reliability of a test is the addition of highly intercorrelated items rather than the mere lengthening of the test. Stephenson (338) questioned the value of knowing the reliability of a test as we now measure it. Factor saturation, he claimed, rather than the reliability coefficient, is the important thing to know concerning a test.

Lindquist and Anderson (315), in a discussion of achievement testing in the social studies, attempted "first, to define as specifically as possible the particular functions which may best be performed by tests of the general achievement type; second, to show how the general achievement test in the social studies must be constructed if it is to perform these functions most effectively; third, to show why a test of this type cannot be made to fulfill other measurement purposes without detracting from its validity with reference to those functions which are distinctive to it; and finally, to present a detailed description of a test of this type and to illustrate concretely its more important characteristics." The authors drew attention to important distinctions between validity of content for inclusion in the course of study, and validity of content for inclusion in a test of *general* achievement, as well as between the concepts of validity as applied to diagnostic and general achievement tests. They also provided concrete illustrations of factors which contribute to high and low validity in individual items in general achievement tests in the social studies.

The fact that there may be a difference between the validity of a test when used to measure status and its validity when used to measure change, the difference between two measures of status, was pointed out by Watson (350). He recommended that good tests, intended to be used "before and after," should furnish coefficients of reliability and validity for change scores as well as for status scores. Worcester (355) suggested the need of research to discover just what type or kind of question is actually asked in the classroom during recitations and in life outside of the school, so that similar questions may be used in tests in order to gain validity. Illustrations of items showing the difference between purely factual and thought, or interpretation, questions are contained in a discussion by Price (327) of tests, objective and otherwise, in the social studies.

Administrative Factors Affecting Test Validity and Reliability

The fallibility of gross statistical measures of test validity and reliability when secured under uncontrolled conditions has been demonstrated by a number of studies which have attempted to measure the influence of certain factors of test administration upon obtained validity and reliability coefficients.

1. *Rate of administration*—Cook (286) and Lindquist and Cook (316) have demonstrated that both the validity and the reliability coefficients for a given body of test materials are dependent largely upon the time in which

these materials are administered, and that unless this time factor is controlled, studies of the comparative validity and reliability of different test materials are likely to prove inconclusive. They defined "optimum administration time" as the length of time of administration of a test in which the highest validity is secured per unit of time and proposed an experimental procedure for determining this optimum time empirically. They investigated the relationship between administration time and validity and reliability for six forms of a spelling test, and showed that the nature of this relationship varied from one type of test to another.

Tinker (342) wondered if the "speed attitude," the feeling of working against time, might lower the validity of tests. He gave two forms of the Army Alpha test to 221 freshmen and sophomores with different time allowances and found that the scores were almost as high and the tests equally valid when the students worked against a time limit large enough to allow all of them to attempt most of the items as when there was no time limit at all. He concluded that it is all right for students to have the speed attitude when taking tests, i. e., to be working under pressure, provided a reasonable time is allowed for the test. Caldwell (282) found very similar results with seventh-grade children on the Stanford Achievement Reading Tests. He analyzed the results on various intelligence levels and found that the students of lower intelligence required more time than the more intelligent pupils to reach their maximum scores.

2. *Practice on tests*—Anastasi (272) conducted an experiment with college students which showed that test reliability increases with practice on the test when the test is administered by the time-limit method and reliability is measured by the odd-even technic. She discussed four factors which sometimes cause variations from the above generalization. Her report includes an exhaustive summation of the experimental evidence on the problem.

3. *Instruction between test administrations*—Copeland (287) explained the data found by three other experimenters, in which tests were less reliable when given at the end of a teaching period than when given at the beginning of the period, on the basis of decreased range of talent at the end of the period due to the instruction received.

4. *Directions on true-false tests*—Weidemann and Newens (351) performed an experiment to find the effect of provision of different sets of test directions for true-false and indeterminate statement tests upon the time required for the administration of the test, upon the reliability of the test, and upon the central tendencies and variabilities of the distributions of test scores. Few significant differences in performance were found between tests with various sets of directions. The final conclusion of the experimenters was that one should "use a set of directions for true-false and indeterminate statement examinations whose definitions for response on the decision scale correspond to the nature of the instruction for each item of the course."

5. *Arrangement of items according to difficulty*—Using 453 pupils in the Minneapolis fifth and eighth grades, Capron (283) found no significant differences in performance due to the arrangement of items in objective tests in spelling, arithmetic problems, and the fundamental arithmetical processes. Easy-to-hard, hard-to-easy, and random arrangements were used.

Measurement of Validity and Reliability

A comparison of the advantages and limitations (due to assumptions involved in their use) of the tetrad technic, the split-halves technic using the Spearman-Brown prophecy formula, and simple correlation of two forms of a test as measures of reliability was made by Dunlap (294). He concluded, among other things, that "the Spearman-Brown formula will give a very close approximation to the reliability of the total form, as split halves will in general be approximately equally reliable." Brownell (279) would be skeptical about this statement, especially if applied to comparatively short, non-standardized tests. He obtained variations from about .30 to .55 in the reliability coefficients of single tests measured by the split-halves method and the Spearman-Brown prophecy formula when the test items were arranged in different orders. Such evidence is taken to indicate the necessity of considering carefully the assumptions underlying this technic when using it to measure reliability. Brownell pointed out that most texts on statistics say very little about the conditions under which the formula may properly be used, and consequently much misuse results.

Handy and Lentz (299) developed a formula to express the reliability of a test in terms of the discriminating value of the individual items, these values, in turn, being computed by comparisons of the responses to each item by the subjects who ranked in the upper 30 percent and in the lower 30 percent on the entire test. Dickey (293) worked out a formula for estimating the reliability coefficient of a test for one range when its reliability coefficient is known for a given range. Because the standard error of estimate of his formula is smaller than that of the similar formula developed by Kelley for the same purpose, Dickey claimed it is the better of the two. Cureton (291) discussed a method and appropriate formulas for finding the reliability of a fallible criterion, such as college success or a series of judgments, against which the validity of a battery of tests is to be computed.

Consistency of Responses

Dewey (292) gave a reading comprehension test composed of several different types of objective items, with several items (of different types) testing each idea, to fifty-five eighth-grade students. The consistency with which answers were made, either right or wrong, ranged from less than 50 percent among the least intelligent students to less than 67 percent among the most intelligent. The conclusion was that a response to a single item or even a single type of item is a questionable measure of a pupil's

understanding of an idea. Similar findings were obtained by Brueckner and Elwell (281) in tests on the multiplication of fractions. They recommended including three or four examples of each kind in a diagnostic classroom test in arithmetic because of the many errors due to factors other than the understanding of the correct procedure required to solve the problem. Brueckner and Hawkinson (280) followed up this study by experimenting to see whether or not placing all four examples of one type together in a row (which procedure would make scoring and analysis much easier than if the related items were scattered throughout the test) would affect the number or kind of errors made. It did not, and so such grouping of similar items is advocated.

Comparative Validity and Reliability of Specific Test Forms

The literature on testing during recent years has contained a decreasing proportion of studies concerned with experimental determination of comparative validities and reliabilities of various types of objective tests. Most of the studies of this character which have appeared, including a number of those which are reported below, are subject to certain general criticisms which may in part account for the decreasing activity in this field of research. These general criticisms are as follows:

First, the studies of comparative validities and reliabilities thus far reported have in most cases attempted to determine the relative effectiveness of the various technics in *general* rather than in relation to specific fields of subjectmatter or in relation to any specific objectives of any given field. Whatever advantages or superiorities any type of test may have, however, are specific advantages in specific situations. Generalized conclusions not only are of little or no positive value but may even be definitely misleading, since they may result in a general condemnation of types of tests which in specific instances or for restricted purposes might be highly valuable. In order to be of value to the test constructor, comparative studies must determine the relative effectiveness of various technics for highly specific purposes. Few studies of this type have yet been made. Because of their highly generalized nature, the majority of studies thus far reported contain little or nothing of value to the test technician in the solution of specific problems.

Second, the majority of reported studies of the relative effectiveness of various testing technics have based their conclusions concerning the "relative effectiveness" of the technics investigated primarily upon determinations of their comparative *reliability* or self-consistency. Objective and dependable descriptions or measures of *validity* are often extremely difficult to secure because of the lack of any acceptable independent criterion of validity. Reliability coefficients, on the other hand, can usually be easily determined. For this reason, there has been a tendency to give undue prominence to the concept of reliability. The reliability or self-consistency of any test or testing technic, however, is of very minor significance in com-

parison to the validity of that techniq in relation to the specific purpose for which it is intended. Considering the fact that the reliabilities of a number of tests intended for the same purpose may, in some instances, even be negatively related to their validities, experimental comparisons concerned primarily or exclusively with the reliability characteristic are likely to contain little of value to the test builder.

Finally, the majority of comparisons between various technics in testing that have been reported have failed to control certain important factors which of themselves could readily account for the differences which have been found. Among the most important of these factors is the skill or ingenuity of the test builder. Investigator A, for example, who is interested in measuring the amount of scientific information acquired by high-school pupils in general science, builds a true-false test and a matching test over the same items of information. In this situation he might show that *his* true-false test is more reliable and valid than *his* matching exercises. This may only prove, however, that A is more ingenious in the construction of true-false tests for this specific purpose than he is in the construction of matching exercises, and may show nothing at all concerning the relative effectiveness with which these technics may be employed by other test constructors in the same or in another situation. The validity of any test in relation to a given purpose is far more a function of the skill or ingenuity evidenced in the application of the technic used than it is of the type of exercise employed. Where this factor of skill or ingenuity in test construction is left uncontrolled, comparisons of obtained measures of validity or reliability of two types of tests may only show the degree to which the test builder has realized or approached the ultimate possibilities of each type of test in the situation involved, and may not indicate at all how they would have compared had their respective possibilities been *fully realized* in that situation. There are few studies in which this factor has been even recognized and practically none in which it has been adequately controlled. Considering, furthermore, the extreme difficulty of controlling this factor in the experimental situation, it seems unlikely that empirical studies will contribute very much to the better evaluation of the various types of objective test exercises. An exception to this generalization may be noted in the case of those few fields where it is possible to provide a detailed and objective description of the manner in which the individual items are constructed, i.e., those fields in which the ingenuity of the test constructor plays a relatively minor role or can be held constant, as in tests of spelling, of the basic arithmetical skills, and of the mechanics of correct writing. Another factor frequently left uncontrolled in experimental comparisons of the type here considered is the factor of administration time, which has been discussed briefly in an earlier reference. The foregoing considerations should be kept in mind by any reader who has occasion to refer to the articles briefly summarized in the following paragraphs.

Cook (286), in a carefully controlled experiment, determined the relative validities of six forms of a spelling test when each was administered at its optimum time. In general, the recall types were found to be superior to recognition tests and the latter to be improved in validity by corrections for guessing.

A five-response multiple-choice test of word meanings, a "same-opposite-neither" test, a matching test, and a multiple-choice test in which the stimulus word was given in a sentence were compared in validity by V. H. Kelley (310) by correlating them with a criterion test in which the subjects were asked to use the words in a sentence. The optimum testing time for each test was determined by experimental procedures and each test given in the time which gave it maximum validity per unit of time. The tests were compared in difficulty and in validity as measured by the percent of agreement between each and the criterion test on individual items. The same-opposite-neither test and the multiple-choice sentence test were lower in validity, though there was not a significant difference between them and the others. The matching and multiple-choice tests were most valid, and the latter most difficult.

Sims (332) experimented with 5-, 10-, and 15-item rearrangement tests on the college level, each item consisting of 5, 10, or 15 headings taken from a psychology text which were to be arranged in logical sequence, and found that they compared favorably in reliability and correlated highly with other measures of achievement in psychology and of intelligence. The 15-item type took longer to write and considerably longer to score but was more reliable than the other two; it was recommended in cases where only a limited number of items was available. Weller and Broom (352) studied the validity of six types of spelling tests. They found that the proof-reading type of test measured something different from what is measured by the recall type of test. The sentence-dictation test was considered most valid. Correlations between types were low; within types, high.

Because he found that individuals were often inconsistent in their answers to items similar in content but different in form, Magill (320) maintained that the burden of proof rests upon those who hold that true-false, multiple-choice, and recall tests are of approximately equal validity and therefore use them indiscriminately. He gave tests of these three types covering the same information to two groups of teachers in service; intercorrelations of resulting scores were high, but individual responses were very inconsistent. Corrections for guessing did not change his results materially. Hurd (308) compared a short-answer with a multiple-choice type of science test in validity, reliability, and difficulty. The tests covered identical subject content. The reliability coefficients computed by the split-halves method were .93 for the short-answer test and .82 for the multiple-choice test. The correlation between them was .78. The latter test was the easier of the two. The short-answer type was said to be more valid because

larger gains were made on it between the first administration of the test (before instruction) and the second administration (after instruction) than were made on the multiple-choice test. (This seems to be a doubtful criterion of validity, especially in view of the difference between the tests in difficulty of items.)

Innovations in test form—The true-false test has been most often "improved." Briggs and Armacost (276) found that oral true-false tests for immediate recall compared very favorably with such tests presented in visual form. The reliability of a 50-item test which they gave to a college class in junior high-school procedures was about .69; stepped up to 100 items, it would be .81. Its obvious advantage is the saving of time and expense involved in mimeographing copies of the test.

An attempt has been made to improve the true-false test by requiring the student not only to recognize a false statement when he sees it but also to know enough about why it is false to be able to correct it. Horn (304) described what she called the "variable answer" test in which the students were to correct the false items; scoring was somewhat subjective. Andrus (273) required his students to correct the false items by crossing out just one word and putting another in its place. Then he built his test so that there could be only one correct one-word correction. It made test-building rather difficult but secured objective scoring and combined the advantages of the true-false, multiple-choice, and completion tests. The test was scored by awarding one point for correctly indicating that an item was false, one for crossing out the right word, and one for putting the correct word in its place. Hevner (301) increased the reliability of a true-false test by allowing the pupil to indicate his confidence in the correctness of his answer and weighting each response according to the degree of confidence shown in it.

The multiple-choice form of test has also been shown to be usable when presented orally. Sims and Knox (333) checked 3-, 4-, and 5-response orally presented tests against a 5-response visually presented test as a criterion, and found them to be a little more difficult and a little less reliable than the visual test but not so much so as to invalidate their use. The 5-response oral test was superior to the other two. Scheidemann (331) suggested asking the pupils to indicate *all* correct responses to a multiple-choice item rather than the *one* correct response. Frutchey (297) explained a practical modification of the multiple-choice type of test for the measurement of ability to apply chemical principles.

Some investigators have reported experiments with test forms widely different from the commoner types. Desiring to measure application of information, Smeltzer (334) compiled tests in which each item consisted of a description of a practical situation and several suggestions as to what should be done by the student in such a situation. Students rated the answers on a scale of one to five. The keys were based on the judgment of several professors, and errors on the part of those taking the test were measured in

terms of their numerical deviations from the key on each response. Brooks (277) described a test used in his history of education and introduction to education classes in which he lists 100 terms and asks the pupils to underline all the terms they can define and then to go back and actually define every fifth one. The percent of correct definitions multiplied by the number underlined gives the score. The procedure saves time; of course, the scoring is subjective to the extent that the terms listed may be variously defined.

Miscellaneous considerations—Lamson (313) raised the question of what happens when a student changes his answer on a true-false test. Her conclusion, based upon 144,000 items from 1,511 papers of college students, was that "it is better to record a second judgment in a true-false examination than not to record it. The chances are two to one that the second judgment will be the correct judgment. It is much safer to change a judgment from true to false than vice versa."

Stalnaker and Stalnaker (337) found that selected distractors (words commonly confused with the word in question) in a 5-choice best-answer vocabulary test were marked more often than chance distractors (those selected at random). There was no interference with discrimination, so their use is recommended.

Item Validity and Reliability

Methods of determining item validity—The importance of measuring the validity, or effectiveness, of a single test item as an aid in the construction of more valid tests has been recognized by investigators, and a variety of methods for computing an index which will give an objective description of the worth of an item have been suggested and used. One of the principal limitations of most of these indexes is that they are concerned only with the correlations between a single criterion and the responses to individual items, and do not take into consideration the intercorrelations between the item responses. Theoretically, the best test is that in which the individual items correlate highly with the criterion but show relatively low intercorrelations. An item with a high index may, therefore, prove less valuable when used together with other highly related items than another item which has only a medium index but which shows low correlations with other items in the test. Another practical limitation of the indexes prepared is that their worth depends upon the validity of the criterion employed, and that reliable independent criteria are not often available. When the criterion employed is the total score on the test itself, the index for an individual item is strictly a measure of the extent to which the item contributes to the reliability rather than to the validity of the whole test. For these reasons, extreme caution must be exercised in the interpretation of results with most of the procedures suggested in the articles reviewed in the following paragraphs.

Lindquist and Cook (316) set up five criteria for an index of discrimination and compared five different indexes with reference to those criteria.

They found the bi-serial r index of discrimination to be the best indication of the value of an item in contributing toward the ranking of students according to general achievement. The calculation of this index requires a great deal of computation, however, and is therefore rather impractical in situations where machines are not available for the purpose. A simpler index which may be calculated without the use of machines and which gives quite satisfactory results was suggested as an alternate procedure.

Richardson and Stalnaker (329) suggested and gave the derivation of a formula for the computation of a bi-serial coefficient of correlation which they claimed was based on more suitable assumptions than the one usually used in analyzing test items. Zubin (357) reviewed three methods of internal validation of test items: the critical ratio, bi-serial r , and association methods. Formulas were worked out to correct for errors due to the customary practice of including the item under analysis in the total score when calculating these indexes. Zubin rated the three methods roughly on the basis of ease of application, limitations, and underlying assumptions.

Long (317) devised an improved overlapping method of measuring item validity which correlated very highly (r of .94) with the bi-serial index on a 110-item general science test. The method depends upon no statistical assumptions, is easy to understand and to compute, is not affected by the difficulty of the item, and has been found to be slightly superior to the bi-serial r index of discrimination in selecting items which correlated highly with the criterion used. Long also developed a further "weighted overlapping formula" which has even higher selective value; it is weighted according to the number of "differentiations" an item makes.

Statistical theory and formulas for the evaluation of test items by the method of successive residuals, which takes into consideration discrimination plus intercorrelation of items so as to avoid having the items all test exactly the same thing, were developed by Horst (307). He also illustrated a work sheet for use in carrying out the mechanical operations involved in evaluating items by this method.

Votaw (349) worked out a method of correcting for guessing when measuring the validity of test items so that comparisons between performances of good and poor students would be measured by the proportions of the two groups who really *knew the answer* rather than by the proportions who *answered it correctly*. The result of the use of this plan in evaluating items was the retention of more items which were selective at the low end of the distribution so that better discrimination among the poorer students was obtained. A scheme was devised for detecting items which had "tricked" good students into making incorrect responses.

Votaw (348) explained a technic for working out a graph with which to determine the selectivity of test items. Arnold (274) presented a variation of this in chart form.

Application of data on item validity—The determination of an index is merely a means to an end. The question of how to make use of it in test construction is an important one upon which there is little experimental evidence. Smith (335) gave a 200-item vocabulary test to 370 bright eighth-grade children, computed the bi-serial r indexes of discrimination of the items, and ranked the items according to these indexes. Then he made four subtests of 20 items each (using the 20 highest ranking items in one subtest, the 20 lowest ranking in another, and two intermediate sets of 20 items, one above the average and one below, for the other two subtests) and used the remaining 120 items as a criterion test. All 200 items were given to two more groups of 500 pupils each. The entire test was more valid when the subtest of lowest ranking items was omitted, but the omission of any of the other subtests lowered the validity of the test as a whole; therefore, Smith concluded that in improving a test the "worst" items should be discarded but all items with indexes above .40 should be retained even though there are many items with much higher indexes. In other words, although these items are of medium low validity as measured by the index used, they measure something which needs to be measured and which the items with higher indexes evidently fail to measure, and so they should not be discarded.

Lindquist and Anderson (315) discussed the factors which affect the discriminating power of items and gave specific illustrations of items which proved ineffective in actual practice, pointing out the reasons for their failure to discriminate well.

Relation between item difficulty and validity—Henry (300) found no significant relation between the difficulty of an item and its validity as measured by the bi-serial r method, the Clark method, the Vincent method, the upper and lower thirds method, and a combination of them all. Horst (305) developed a formula to express the difficulty of a multiple-choice test item as a deviation from the mean of the group in standard deviation units. He deduced the idea that multiple-choice items with fewer alternative responses which are equal in difficulty are more valid than those with a larger number of choices presenting a wide range of difficulty.

Reliability of test items—Holzinger (302) worked out formulas giving the standard error of response or measurement of a single test item. He stated:

The response error of series of items is already known. The problem, then, is to find this error for a single item and show the relation to the standard error of a series of such items. It is hoped that these new formulae will be of some value in building up tests of items of known reliability and in predicting the final reliability of a number of such items when combined. Conversely, it is believed the formulae may be useful in appraising the reliability of tests of different lengths by reducing the measure of reliability to the average item basis.

Collection of data for test validation—Horst (306) described a procedure for building a test of personality and mental alertness in such a way as to try out a large number of items with a minimum of test administra-

tion. He divided his try-out group into two groups, gave each a set of items of two types, and then predicted what the score of each individual would have been had he taken a test on all the items of one type. In this way twice as many items may be tested as is usually done. The method and formulas may be applied to any type of test.

Scoring

1. *Errors in scoring*—Rauth (328) compiled data showing that errors are frequently made in scoring objective tests, and listed the most common types of errors made in scoring various test forms.

2. *Methods of scoring common forms of objective tests*—The use of mechanical devices and scoring forms in an effort to save time and money and to be completely objective represents the trend in scoring procedure. Bawden (275), Cuff (290), Fay and Middleton (296), and Pressey (326) have reported the use of scoring forms, separate sheets of paper ruled in various ways, upon which the pupils indicate their answers. This plan permits the same test papers to be used again and again. Such score forms may be used with any of the more common types of objective tests; ingenious suggestions for scoring them are many. Cuff (290) ran the answer sheets through a mimeograph so arranged as to encircle the correct responses and make speedy tabulation of the results possible. Fay and Middleton (296) and Pressey (326) recommended a cardboard stencil key. Pressey also described a machine procedure which secured speedy and accurate results—a part of what he called “the coming ‘Industrial Revolution’ in education.”

Cuff (289) told of a mechanical device for scoring multiple-choice test items by *weight*, which is from 10 to 40 times faster than scoring by hand and very much more accurate—practically perfect. Other forms of objective tests undoubtedly could be scored similarly.

3. *Methods of scoring less common forms of tests*—Elderton (295) presented objective methods for the scoring of maps drawn by students and for the combining of time intervals and errors into a single score on tests of mental imagery.

Brown, Bartelme, and Cox (278) explained and advocated the use of a procedure for scoring tests in which items have been so scaled that the scale *values* of the items missed and passed can be taken into consideration. The score of the individual is taken as that point on the scale where the average deviation of the correctly marked items above it equals the average deviation of the incorrectly marked items below it. Examples are given to indicate the superiority of this method over the one commonly used by Thurstone and others, which involves the scale positions rather than the scale values of the items.

T. L. Kelley (309) developed a formula to determine the weight to assign to a given response in a test designed to compare a person's interests with those of some homogeneous group or class of people.

A special 68-item examination devised to measure the ability to evaluate two types of objective test items was given to a University of Arkansas class in construction and evaluation of objective test items by Gerberich (298). The examination was scored by four methods. The optimum method, as determined by its reliability and correlation with a final examination and term grades, involved a weighting of one point for each "good" item so designated and two points for each "poor" item so designated.

4. *Corrections for guessing*—Melbo (321) analyzed the results of a 50-item true-false test given to 1,480 high-school and college students and decided that the right-minus-wrong formula to correct for guessing was fallacious—that it was better to consider the number of correct responses as the score. Such results would have been expected in view of a study by Kruege (312), who made an empirical check on the laws of chance and showed that the right-minus-wrong formula provides a fairly valid correction for guessing with tests of several hundred items but that in tests of less than 100 items many spuriously high and low scores will be obtained in spite of the use of a correction formula.

Zubin (356) developed formulas to correct for guessing in determining the mean and the standard deviation of a matching test. They are quite simple and easy to use. They may be applied to individual scores provided the number of questions is sufficiently large.

The Influence of Objective Examinations

1. *Upon instruction*—The March, 1935, issue of the *Journal of Educational Research* (354) contains the opinions of a number of leading testing authorities concerning the effects of measurement upon instruction. The general opinion seems to be that the influence of measurement upon instruction may be either beneficial or harmful depending primarily upon the nature and quality of the measuring instruments used.

2. *Upon learning*—The so-called negative suggestion effect of true-false items has been a subject of much research. Keys (311) found slightly harmful suggestion effects from false items in an educational psychology test. McClusky (319) showed clearly that college students miss false items more often than true ones and compiled data from which he concluded that there were detrimental negative suggestion effects from true-false tests, though he recognized that, in any case, they were temporary in nature. Sproule (336) carried on experiments at the fifth-, seventh-, and ninth-grade levels and concluded that "the present evidence does not justify a condemnation of the true-false test on the basis of false impressions that it produces." He found that allowing students to correct their true-false tests offset practically all negative effects and contributed to positive learning, so he decided that it was safe to use such tests as low as the fifth-grade level if students were allowed to correct their papers. Similar decisions were made by Ross and Pirie (330) after experimenting with college students. They found that true-false tests did not inculcate false impressions when the correct answers were given to the students after the

test. Thus the evidence would seem to indicate that the true-false test, *when properly used*, is a valuable instrument for instructional purposes.

Meyer (322) conducted an experiment which showed that when pupils study for a recall test they retain the knowledge longer than when they study for a recognition test, so that it is best not to let them know ahead of time that they will be given a recognition type of test. McClusky (318) illustrated the value of correcting tests in class and of discussion at that time as an aid to maintenance of achievement.

3. *Upon study habits*—Studies by Meyer (323), Class (285), and Terry (339) have shown clearly that students study differently in preparation for different types of examinations, a fact which should be known and considered carefully by instructors when deciding upon their testing procedures. Students study for details when preparing for objective examinations; they study the larger aspects of a subject—to get a general picture of the material—when an essay test is announced. Meyer (323) found that in studying history they draw maps and make summaries more when studying for an essay test and do less underlining, and that individuals studying for completion tests study harder and make out sample test questions more often than those studying for recognition tests. All of these studies have been made with college students; the generalizations may not apply to pupils at lower grade levels. Crawford (288) has compiled a list of suggestions on how to study for objective tests in general and for true-false, completion, and matching tests in particular.

Bibliographies on Test Construction

Lee and Symonds (314) have written an excellent summary of investigations concerning objective tests reported between October, 1931, and October, 1933; they included a bibliography of 104 references. Holzinger and Swineford (303) have compiled three annual annotated bibliographies of selected references on the theory of test construction.

General Studies

The Committee on Educational Research of the University of Minnesota (324) published an account of the experience of that institution with the construction and use of the examinations employed in the new General College and in other departments of the university. The chapters by Alvin C. Eurich, Edgar B. Wesley and Renata R. Wasson, Henry Kronenberg and Edgar B. Wesley, Alvin C. Eurich and Francis S. Appel, and Clara M. Brown particularly contain good suggestions for the construction of objective test items. The appendix contains an interesting collection of sample test items from the Minnesota examinations, which should provide valuable suggestions to the test builder.

A similar and more exhaustive and varied collection of sample test items was prepared by the Board of Examinations (284) of the University of Chicago. The latter collection is prefaced by a brief introduction containing suggestions for the construction of objective test materials.

CHAPTER IV

Recent Developments in the Written Essay Examination

FROM 1923 to 1935, research studies on the essay test may be grouped as follows:

1. The reliability and validity of essay marking, including comparisons with objective tests
2. The measurement of the same or different mental functions tested by essay and objective tests
3. Student preparation for and reaction to essay and objective testing programs.

Reliability and Validity

Wood (389) pointed out that the new type (in this case the true-false) is more reliable when compared with the old type essay examination, and further stated that the essay type "is most apt to measure: cogency of expression, organizing acumen, and reasoning ability." In a second report, he emphasized that "the essay examination is best adapted to the measurement of critical capacity and reasoning ability. . . . The best essay examination is the one which allows the student to choose two questions out of five. . . . The essay law examination is indispensable . . . and . . . it can be improved."

Sims (381), using "the distinctly good questions" collected by Monroe and reported by Odell, found that 34.5 percent were simple recall, 35 percent were short answer, and that only 30.5 percent were discussion questions. Sims stated that the first two are definitely objective in nature and should be built according to established principles of objective test construction, while the third type is more subjective. He suggested that a more satisfactory method of marking the subjective type should be developed. In another paper he (384) indicated how to reduce by one-half the range and probable error of the marks given by different graders. He considered the scores given on essay examinations as raw scores and converted them into the particular grading system of the school by setting the passing point at 1.5 standard deviation of the raw score. Using this procedure, he (383) compared two forms of an essay test of ten questions each with two forms of an objective test each consisting of 34 completion and 40 true-false questions which were given to 80 students in general psychology. The average coefficient of objectivity of rating the essays was .77, their reliability was .72, and the correlation between the essay and objective tests was .70.

Weaver and Traxler (387) based their study upon 5 objective and 5 essay tests made upon 2 units of history and given to 38 pupils. They

found correlations from .30 to .60 between the tests and a combined 4-essay test as a criterion and again with a combined 4-objective test as a criterion. Separate test pair reliabilities were not computed. The essay type question used by Weaver and Traxler would be classified by Sims as calling for an objective short answer.

The viewpoint of Sims must also be borne in mind when considering Leighton's suggestion (371) for reducing the sampling error of the essay test by the addition of a large number of short answer questions. He found that when factors to be measured are carefully determined before constructing the essay examination, when the questions are planned and clearly stated, and when criteria for judgments are set up, there is great increase in the agreement of scorers' grades. With these criteria, two graders using the subjective method of grading papers as a unit were able to correlate .63 when grading very subjective material like that of philosophy. They used a 1, 2, 3, 4, 5, fail, ranking of papers. Graders of biology, when using the more objective method of grading each question separately, correlated between .63 and .90 in their judgment of the papers.

Tharp (386) compared old and new type foreign language grammar tests with a third semester test consisting of both essay and objective questions. He found a correlation of .90 for the objective and .83 for the essay test when compared with the grades given in the course as determined by the final examination. He suggested that the objective test will measure more accurately than the essay test alone.

Eells (366) had 61 teachers grade the same essay test material at the beginning and at the end of an eleven-week interval. He found the variability of judgment between the same individual to be as great as the variability between different individuals. Correlations ranged from .25 to .51. The material used consisted of the geography questions from Ruch's experiment and the history material from a previous experiment of Paulu.

Over a period of two semesters, Peters and Martz (379) made an elaborate study of short true-false, multiple-choice, completion, and essay tests in elementary and secondary subjects and correlated them with the final grade given in grades 2 to 12 inclusive, involving 252 students. They concluded from the 196 correlations computed that completion and essay tests do not vary greatly in validity where the criterion of final grade is used. The true-false test was slightly less valid than the completion and essay tests, especially in the elementary grades. The multiple-choice and essay-discussion tests were equally valid in the elementary school, but in high school the essay was more valid than the multiple-choice test. The criterion of final grade with which each of the ten-minute experimental tests was compared was determined one-third by teacher-made objective tests and the ones mentioned above. The latter actually contributed about one-twelfth of the criterion.

McKee (372) equated two groups of 50 students for intelligence in elementary and advanced freshman English and compared their semester

grades, determined for the most part by a subjective estimate of written work, with a prognosis made first by an objective test and with a second prognosis made by the use of a theme. Both the test and theme were written at the beginning of the semester. He found the objective test to be a far more reliable prognosticator of superior students in freshman English than the written theme.

Odell (376), in an extensive study involving 23,500 pupil answers to thought questions scored by 57 raters, concluded that the reliability of ratings given with scales was not significantly higher than that given without scales. He further found that with raters having no teaching experience the scales tended to assist in fixing the general test standards but they did not increase the reliability of the actual rating of single answers. This finding is similar to that of Ruch and Stoddard (380) but opposite to that of Hudelson (368).

Monroe and Souders (375) compared two sets of examination papers prepared by pairs of teachers working together and again working individually both in the construction and the grading of final examinations. This work, carried on by 1,736 high-school teachers, showed that while in the best standardized tests both constant and variable errors are distinctly less than corresponding errors in examination grades, nevertheless written examinations may be so improved that differences in accuracy of examination grades and scores may not only be lessened but may yield about the reliability of .65 as found in many widely used standardized tests. They further believe that their results contradict those of Elliott and Starch because the latter failed to differentiate between variable errors (those errors due to different educational aims, different standards of excellence, etc.) and constant errors (those due to no distinction between scores and grades).

Cochran and Weidemann (361), by setting up standard procedures for grading "explain" and "discuss" essay test questions, found consistency coefficients ranging from .78 to .98 between experienced scorers in history. Thirty-five teachers with experience in several high-school subjects and with only ten-minute training in the standard procedures of scoring were able to produce an average consistency coefficient between their first and second scoring of .78 and an average objectivity coefficient, based on 237 correlations of .56.

Leighton (371) criticized the work of Ruch, Elliott, and Starch in two ways. First, the experiment with a single paper and several judges giving a numerical grade is artificial. He suggested that a more significant experiment would be the ratings of a group of papers using the usual comparative value grading rather than the percent technic. Second, there is no evidence that a definite criterion for judgment was offered for the essay test, though such a criterion is deemed extremely important by these men when building an objective test.

Kinney and Eurich (369) summarized the research on comparison of different types of tests saying that there are few conclusive results. They pointed out a double need: first, for experiments more coordinated and of a wider scope than the usual "controlled experiment;" and second, the need for pooling administrative experiences with regard to test construction and use.

The Measurement of Mental Functions

Corey (362) gave an essay test consisting of 6 questions, and an objective test consisting of 96 multiple-choice questions and 13 matching questions to 102 students in educational psychology. The correlation between the essay and the objective tests when corrected for attenuation was .93, thus indicating that the two tests measured the same function. The results of the Army Alpha test for the above group, when correlated with the essay test and corrected for attenuation, yielded a coefficient of .39; but when the Army Alpha was compared with the objective test and also corrected for attenuation the coefficient was .62. Corey concluded that when the essay and objective tests cover the same subject the latter is more closely related to the Army Alpha than is the essay test. The results suggest the possibility that other factors of this study may require a higher degree of control.

Paterson (378) gave a one-hour essay test and a one-hour objective test to students in his five-hour two-quarter course over a period of two years. He found the correlation between the old and new type tests to be .52. As the correlation between these two tests were as high (.52) as the validity of the lowest (essay) test, he concluded that the two examinations were measuring the same thing. It should be noted that the studies by Corey and Paterson compared unimproved essay tests with tests consisting of two or more forms of improved objective tests.

Sims (382) gave thirty-three students an essay examination of six recall and four discussion questions. Eight readers scored the recall questions with a key and scored the discussion questions by sorting them into normal distribution groups. All papers in each section of the distribution were marked with a predetermined score. The variation among the readers was slight, the objectivity of the examination was .97, and the reliability of the examination was .84. The recall questions and the discussion questions did not seem to measure the same thing as the correlation between them when corrected for attenuation was only .53. When compared with an objective test taken by the same group the recall questions yielded a correlation of 1.01 when corrected for attenuation, while the discussion questions yielded a coefficient of .64 when similarly corrected, thus indicating that the latter type of question measured something quite different from that measured by the objective test.

Cochran and Weidemann (360) conducted an experiment consisting of 4 to 8 improved essay questions making up an "explain" essay test, and a second test consisting of from 50 to 100 simple fact answer questions over

the same material covered by the essay test. A set of 4 tests was constructed to cover the work of one semester of American history. The tests were given every 8 weeks over a period of four consecutive semesters. The improved essay test was given first and the simple fact answer test was always given on the following day. The median consistency coefficient for the 4 simple fact answer tests was .87; the median consistency for the 4 "explain" essay tests was .55; while the median community of function value was .59 with a range from .54 to .82. As a result of the use of supplementary statistical technics it appeared that the overlapping of mental functions measured by the "explain" essay and the simple fact answer tests under actual classroom conditions was about 60 percent; that about 40 percent of the mental functions measured by the former were not measured by the latter; and 40 percent of the mental functions measured by the latter were not measured by the former. Similar experiments with the "discuss" essay and simple word answer fact tests were conducted by the same authors (359). The results resembled those for the "explain" essay and simple word answer tests.

Weidemann and Newens (388) experimented with the "compare and contrast" essay and true-false tests covering the same material. Each of the seven units of test materials was based upon the content of the course during the two weeks immediately preceding the administration of the given test unit. The consistency of the essay was found to be as high as that of the true-false test. Under classroom conditions the two tests overlapped in mental functions to the extent of from 50 to 70 percent in terms of median values. Approximately 30 to 40 percent of the mental functions measured by the "compare and contrast" essay tests were like 30 to 40 percent of mental functions measured by the true-false tests. The evidence on overlapping of mental functions measured by essay and objective tests is insufficient to warrant any definite conclusion.

The foregoing studies raise the question of what is the nature of the mental functions which such tests measure.

Student Preparation for and Reaction to Testing Programs

Crawford and Raynaldo (364) made twenty comparisons in fourteen different classes. One-half of the students in each class were required to take notes on the lecture while the other half just listened. The next day the groups were reversed. True-false and essay tests were given after each of four such rotations. Those who took notes were inferior as measured by the true-false test but superior as measured by the traditional essay test. The authors suggested that the true-false test emphasizes unorganized use of fact material. However it must be noted that student note-taking does not necessitate organization of material.

From a questionnaire, Douglass and Tallmadge (365) found that students preparing for an essay examination said that they read and reviewed generalities and trends, attempted to draw several important conclusions from tables, formulated personal opinions, and read notes

on the text and lectures carefully but without picking out details to be memorized. Students stated that when they prepared for objective tests they learned tables and minute details of materials covered and tried to remember the exact words of the book and other specific points. The procedure of learning was given by about an equal number of students in studying either for the essay or objective test.

Meyer (373), upon examination of studies by Terry (385), Douglass and Tallmadge (365), and Crawford (363), concluded that essay examinations result in a superior type of preparation, more adequate learning, and more recall than the true-false, multiple-choice, or completion types of examination. For the most part, these results are based upon questionnaires.

Meyer (374) divided 124 students into four groups, each studying for three two-hour periods on Civil War history. One group was told to study for a true-false test; a second group studied for a multiple-choice test; a third group studied for a completion test; and a fourth group studied for an essay test. At the end of the study periods all groups were given all four types of test. Five weeks later the same testing procedure was again followed. Meyer concluded that the examination set which the individual has is of fundamental importance in learning and retaining sense material. The essay examination set should be used in preference to any objective examination set if the student is to recall material in an organized fashion and to know facts when cues are not given.

In the field of student opinion Klise (370) found that of 2,065 students at Iowa State College 79 percent preferred the true-false, multiple-choice, and completion combination of objective tests; 18 percent preferred the essay test; and 3 percent were undecided.

Hastings (367), in experimenting with examinations constructed by students in college English, concluded that such procedure leads to student review work, increases cooperation between instructor and student, and results in an examination which is satisfactory to the instructor.

Summary

Research with the essay test has, in the main, been conducted with the traditional or unimproved form and comparisons have been made with improved forms of objective tests. Such comparisons indicate that the improved objective test measures as good as or better than the unimproved essay test. When analysis of essay tests into defined types such as *list*, *outline*, *describe*, *compare*, *contrast*, *explain*, *discuss*, *develop*, *evaluate*, and *summarize* is achieved for instructional use, comparative studies using improved essay types and improved objective types of tests will become possible.

The use of both objective and essay tests seems to be a better basis to evaluate the achievement of such varied results as specific, general, and organized information.

The foregoing studies give a slight suggestion of two developing tendencies:

1. That objective tests are used to measure certain phases of achievement in such subjects as mathematics, physics, and biology; while both objective tests and essay tests are useful in measuring certain phases of achievement in such subjects as history, social sciences, and philosophy.

2. That objective tests are useful on the elementary-school level; while both objective and essay tests are useful on the college level.

CHAPTER V

Achievement Tests in Colleges and Universities

Factors Determining Significant Characteristics of Educational Tests

THE CHARACTERISTICS significant for an achievement test are determined in any particular case by the uses to which the test is to be put and the effects resulting from such uses. Obviously, tests should be so constructed and used as to promote rather than hinder important educational values. Hence, studies of the uses and effects of tests are essential to the sound development of educational testing.

In the college testing field there have been few carefully conducted studies demonstrating potential uses of achievement tests. Investigations of desirable and undesirable effects of tests are still more limited. Most publications on this topic are arguments for or against various uses, or else present very incomplete and inconclusive evidence of values promoted or hindered by educational testing.

Uses of Achievement Tests in Colleges

In the selection and guidance of college students, achievement tests as well as aptitude tests are frequently used. Nevertheless, few studies have been made to evaluate this practice. The Committee on Personnel Methods of the American Council on Education (390), after reviewing studies showing the unreliability of teachers' marks, of Regents' Examinations, and of College Entrance Board Examinations, proposed to develop many forms of comparable tests in the various subject fields on the ground that these tests would prove more effective bases for selection and guidance of students. No evidence for this claim was presented. The Study of the Relation of Secondary and Higher Education in Pennsylvania (425) demonstrated wide variability in the functions tested among students within the same college, among the average scores of the various colleges, and among various prevocational groups within the colleges. This investigation suggested but did not establish the possible effectiveness of achievement tests in educational guidance. The annual reports of the American Council's Committee on Educational Testing (418, 419) gave similar data on variability and also suggested the use of tests in guidance. B. D. Wood (459) and E. P. Wood (460) maintained that the major function of testing is guidance and that this function was best met by cumulative records of continued testing. Crawford (402), however, presented data from his work with Yale students which clearly questioned the effectiveness of present achievement tests in the guidance of students. He also doubted the comparability of most tests.

Wagner and Strabel (456), in studying the predictability of foreign language marks at the University of Buffalo, found that marks in this field could be more exactly predicted than in any other field at that University. The average of the Regents' grades in language gave the highest predictive index, whereas the cooperative French test did not foretell success as well as high-school marks. Reeder (438) found a low relation between placement tests and success. Since the criteria of success were the marks given by college instructors, the poor showing of certain tests in prediction may be a fault of the marks rather than the tests. Palmer (436), on the other hand, found a correlation of .70 between pre-study and post-study scores when using the cooperative physics test. He suggested that such test scores made possible great improvements in guidance. In support of this position, Terry (447) found that the Van Wagenen Reading Test administered at the beginning of a course in educational psychology gave a correlation of .72 with the final course examinations.

With reference to the use of achievement tests in the placement of college students in the foreign languages, Gausewitz (408), Henmon (414), and R. E. Monroe (432) collected evidence indicating the value of this procedure. Brigham (397) reported the new College Entrance Examination Board plan for combining entrance testing with placement testing. He presented data indicating the greatly increased reliability of the College Board Examinations in English. Fletcher (407) presented no data but described the use of examinations in giving college credit for post-graduate work in high school and for out-of-school educational progress. Stalnaker and Richardson (446) defined carefully the criteria demanded of tests used for giving scholarships to college students.

In determining the characteristics of college students, Betts (393) used a test on contemporary affairs which he constructed and administered to graduate students in education at Northwestern University. Ellis (405) used the Iowa High School Academic Contest Test in World History with University of Missouri freshmen to discover their achievement in comparison with high-school students.

Jones (420) found sixty-six colleges using a comprehensive examination for the degree in two or more departments. Eighty-five colleges were using this type of examination in at least one department. From this investigation, he contended that the major function of the comprehensive examination is to test the student's facility in bringing to bear upon an exercise ideas and facts from several subject fields.

The most frequently reported use of achievement tests in colleges is in studying the results of various types of educational experiences, procedures, or materials. Haggerty (409), in the North Central Association's study of standards for higher institutions, used the results of tests as one index of the quality of the college. Cheydleur (401) employed foreign language tests to evaluate the success of an experiment in teaching French to adults. H. T. Tyler (452) determined the effectiveness of a remedial read-

ing course by using the Nelson-Denny and the Iowa reading tests. D. F. Miller (431), and Price and J. A. Miller (437), by the use of tests, determined the effectiveness of certain methods of segregating students in zoology courses at Ohio State University. Meyer (430) appraised remedial procedures in zoology by similar tests. Bowden (396) employed the Allport-Vernon Test of Social Values in judging the effectiveness of a course in social psychology. Sinclair and Tolman (441) employed tests to estimate the relative effects upon open-mindedness of an arts college course and a technical course. Hartmann and Barrick (411) attempted to use the Carnegie General Culture Test to determine the effect of college education upon general cultural information. Welborn (457) tried to use tests to determine the nature of logical learning in college classes.

As a means of measuring teaching efficiency both Barr (391) and Hartmann (412) utilized tests. Both presented evidence of the complexity and difficulty of the task.

The studies during the past three years have extended the uses of achievement tests in college. Formerly, their uses were usually restricted to marking students. More and more they have been used in selecting and guiding students, in predicting college success, in placing students within courses and within sections of the same course, in assigning advanced college credit on the basis of examination, in awarding scholarships, in studying the characteristics of college students, in granting diplomas or honors, in studying the effectiveness of educational procedures, and in evaluating teaching ability. The variety of uses is a chief reason for the barrage of criticism which has been directed against many of the tests constructed during the past ten years. The commonly accepted technics of test construction do not produce tests appropriate for such varied usage.

Values Promoted and Hindered by Testing

The effects of testing are undoubtedly significant and varied. They are probably dependent upon the nature of the test, the procedures employed in giving the test, the use to which the test results are put, and the traditions of the college as well as the attitude of the individual student. If testing is a potential power for educational good or evil, it is most unfortunate that we do not have thorough studies of the effects of testing under varying conditions. Most publications dealing with this topic present points of view rather than data.

Douglass (403, 404) pointed out what he believed to be crucial dangers in the testing movement, namely, the narrowing of educational purposes, materials, and procedures because of the failure of tests to cover the broader outcomes of education. He emphasized, as another danger, the stimulation of students to frantic but not well-directed study. Similarly, Krey (423) believed that testing which did not closely parallel the objectives of instruction was potentially harmful. W. S. Monroe (433) main-

tained that test makers had now become the real curriculum makers because of the effect of tests in directing the efforts of pupil and teacher.

Counter claims were also advanced. Hanford (410) reported that the comprehensive examinations at Harvard had a beneficial influence on instruction by emphasizing the interrelation of courses. He also maintained that the examinations stimulated students toward definite and desirable goals. Boucher (395) found that the comprehensive examination program at the University of Chicago had forced instructors to formulate their objectives and clarify their purposes. Lowell (429) considered tests to have important potentialities as teaching devices. Kulp (424), using weekly tests with graduate students, found an increase in the amount of learning under these conditions. Keeler (421) maintained that educational measurement when it goes hand and hand with instruction stimulates both teachers and students to better efforts. Theisen (449) stated that the effects of measurement thus far have been to turn educational efforts toward reconstruction of curriculums and procedures rather than to continue the status quo. Barr (391) claimed that continued educational measurement is necessary for the improvement of education and that testing is not in itself harmful. Evils come from poor tests or their unintelligent use.

Lindquist and Anderson (427) attempted to meet the criticism of testing by distinguishing two types of tests, the general achievement test and the diagnostic test. In discussing the general achievement test, they advocated restricting the test to what has been learned. They would include a large proportion of items which test only for understanding at a low level and they oppose measurement of desirable emotionalized attitudes. They believe that the dangers inherent in such narrow testing may be avoided by careful administrative procedures. Richardson and Stalnaker (439) claimed that an achievement test need have no pedagogic value and ought not to measure effective qualities. Wilson (458), on the contrary, recognized the influence of testing and advocated a plan of administration by the teachers themselves of all tests which affect teaching. Woody (461) also admitted the possibility of evil effects from testing but maintained that the danger represented a challenge to improve the testing movement.

This brief summary of conflicting opinions shows clearly the need for thorough investigations of effects of various types of tests under varying conditions. If testing may be a power for both good and evil, we need to know more about the values which may be promoted and hindered by testing and the characteristics of tests which will facilitate rather than hinder desirable learning.

Philosophy of College Testing

The major change in the field of measurement in recent years has been the development of a broad philosophy of evaluation. Previously, the important issue in test construction was the relative merits of the various forms of objective test devices. Should one use a true-false test or a

multiple-choice test, a multiple-choice test or a matching test, a completion test or a true-false test? Attention was directed to the type of test device which could easily be used for particular kinds of subjectmatter. At the present time, attention is focused on the kinds of evidence which indicate the attainment of various important outcomes of teaching. College courses are offered in order to facilitate certain desirable changes in students. An achievement test is a means of discovering the degree to which these desired changes in students are taking place. The nature of the test will vary with the nature of the changes sought. Evidence of the recall of information is different from evidence of the use of information. Evidence of enjoyment of literature is different from evidence of the ability to judge literature. As test makers have come to recognize that they are seeking various kinds of evidence, they have come to realize that varied test procedures are necessary for collecting the evidence. Educational purposes or objectives have become increasingly the basis upon which tests are constructed. Hence, a first problem of the test maker is to formulate the teaching purposes.

Formulating the Objectives

Although objectives constitute a curriculum problem, they are also basic in developing examinations. When objectives have been formulated in designing a curriculum, they are frequently neglected by the test maker. Examinations are devised without reference to the purposes of the course or of the college. Boucher (394) pointed out that not only the course content and methods of instruction but also the evaluating instruments composing comprehensive examinations must be determined by the objectives set up for courses.

Johnson (417) stated that the important preliminary step in constructing tests in science courses is to formulate the objectives of the courses. According to Kelley and Krey (422), any examination designed to test learning in the social sciences must consider the objectives of teaching in this field. Uhl (455) emphasized the importance of appraising heretofore neglected outcomes of teaching. The need for getting evidences of changes in character and attitudes as well as changes in intellectual accomplishment was pointed out by Robbins (440).

Teachers often assume that if a student has acquired much of the information in the course, he has also reached other important objectives to an equal degree. Information tests are used as the only measure of achievement. Eurich (406) gave evidence that in the achievement of objectives in freshman English many relationships are very low and none are high. Similar findings were obtained in other fields. Hence, in order to get a clear picture of the growth of students, it is important to get evidence of changes in the direction of each of the important objectives.

In the absence of objectives based on curriculum studies, Tyler (453) suggested two methods useful in formulating the objectives. One method

was to consider the general purpose of a course and to analyze it into its subfunctions. The analysis continues to the point where the objectives become clear and useful in teaching and appraisal. A second method was to consider the course content, the teaching and administrative procedures, and with reference to each topic and each procedure ask: Why is this a part of the course? What change in student behavior is this expected to bring about?

Clarifying the Objectives

Usually after objectives for courses have been formulated, little has been done to clarify the objectives in terms of student behavior. Unfortunately, the objectives remain vague and nebulous statements. They do not describe the kind of behavior representing evidences of changes in the direction of the objectives. Hence, in constructing examinations, the objectives have often been passed by. Jumps have been made to a test exercise. Research workers in the testing field have frequently skipped the questions, "Does the kind of behavior required on this test express an important objective of the course? Do the kinds of behavior required in this test give evidence of all the important objectives of the course?" These are the significant problems when establishing the validity of tests.

In the field of literature, Carroll (399) defined prose appreciation as the ability to distinguish between literary selections previously judged to be good or poor. This definition became the basis for his test. The significance of definition is seen by the fact that the test requires of the student, kinds of behavior which are different from other kinds of behavior which are often called appreciation. Noll (435) maintained that when students mark certain carefully chosen statements as true or false or doubtful, the students are exhibiting evidences of scientific thinking, consisting of habits of accuracy, suspended judgment, open-mindedness, intellectual honesty, criticalness, and the habit of looking for true cause and effect relationships. The question arises, Is the kind of behavior involved in identifying true and false statements, evidence of scientific thinking? Turnbull and Griffith (451) reported some objectives of college home economics courses and the kind of behavior describing each of these objectives.

Collecting Test Situations

A third problem in preparing an examination is that of collecting the kinds of situations in which the behavior is expected to take place. This aspect is usually neglected. It has been tacitly assumed that a paper and pencil situation is the kind of situation in which the behavior is expected to take place. If we wish to find out how well a student uses arithmetic in situations which he meets in life, his behavior in those kinds of situations should be noted and recorded. If we wish to find out how skilful a student is in the chemistry laboratory, we should observe his behavior in the

laboratory as he works on a variety of problems. The examination would consist of situations which would be representative of the variety of situations in which the student has an opportunity to use arithmetic or to work in the laboratory.

What are the conditions under which the behavior is expected? Are the students expected to remember the information pertinent in answering the questions in an examination or may they have the opportunity to consult sources of information? Stalnaker and Stalnaker (443, 444) pointed out the advantages of and results obtained in an open-book examination. There is need to investigate in similar fashion other important conditions characterizing test situations.

A second aspect of the problem of collecting test situations is that of the size of the sample. As the sample of situations increases in size, the results obtained are more stable. The test is said to be more "reliable." Lindquist and Cook (428) presented evidence showing that the reliability and validity coefficients of certain spelling tests are a function of the time limits in which the tests are administered. The time limits were based upon the percent of students finishing each test. This is a further illustration of the conditions of the test situation. How much time should be allowed for the test? Should the time limit be that necessary for 25 percent of the pupils to finish, or should it be long enough for all pupils to finish? These questions go back to the meaning of the objective. If the test is designed to collect indirect evidence, the optimum time limit is one which gives the best index of the direct evidence.

Recording the Behavior

Two chief uses of a record of the student's behavior are that the behavior may be evaluated at a later time and independently by other individuals. On a paper and pencil examination, the student makes his own record. It is often assumed that this is the only kind of record of changes in the behavior of students. Charters (400) reported the use of the anecdotal record at Rochester Athenaeum and Mechanics Institute. Significant observations of students were recorded. These descriptions of behavior could be evaluated independently by others at a later time. Hence, it is important in recording the behavior to separate the descriptions of the behavior and the evaluations. In an anecdotal record the evidence of behavior is available. In a trait rating device, the evaluation only is recorded.

A record may also be the product resulting from student behavior. Students' themes and art productions are evidence of changes in behavior. A frog dissected by a student is evidence of the student's ability to dissect a frog. R. W. Tyler (453) reported the use of a checklist in recording the behavior of a student in using a microscope and in describing the student's mount. Tharp (448) used a phonograph in recording students' French pronunciations. The records were evaluated at later times by other individ-

uals. Slow motion pictures have been used as diagnostic instruments in recording the behavior of athletes and getting recorded evidence of close decisions. Likewise motion pictures may be used in recording the behavior of a speaker. Photographs may be used as evidences of physical growth.

Evaluating the Behavior

Most standardized tests yield a score for a student *in a subject*. The interpretation of the result is in terms of the relative size of the score in the subject. Recently there has been a shift in interpretation, from achievement in each subject to achievement in each of the important objectives of a subject or in an important objective of several subjects. From this angle the evaluation of a student's achievement can be made in light of the kind of behavior expressing each objective, and not only in terms of a score in a subject.

The essay examination received much criticism due to the variability among readers in grading the same examination paper. In grading essay examinations, it is assumed that readers are evaluating the student's reactions on the same objective. This assumption is not always sound. For example, in grading a test on ability to use information in new situations, some readers forget the objective and grade the test on the amount of information the student has stated instead of the degree to which pertinent information is used in solving a new problem. Wide variations in the scores result.

Trimble (450) studied the objectivity of oral examinations. The students' oral responses were evaluated on eight traits by three judges. The interrelationship between the ratings of the judges on the traits ranged from $-.23$ to $.80$.

Another assumption involved in evaluating the behavior on an essay question is that the readers have the same standard of response in mind. This assumption can be checked in a meeting of the readers at which time the ideas which are expected in the responses are brought together and the numerical values allotted to each idea may be decided upon. Each reader may be given a copy of these specifications to follow in grading the essay questions. Stalnaker and Stalnaker (445) reported that when readers "analyze the ideal answer," and assign "a certain number of points to each significant part of it," the readers agree more closely. Agreement in evaluating English examinations at the University of Chicago increased from $.42$ to $.92$. Horney (415) found close agreement in scoring a test of understanding chemical principles when the responses were evaluated with a definite type of behavior as the standard.

A prevalent misconception of an objective test was clearly discussed by Brownell (398). The misconception assumes that an objective test lacks subjectivity and hence is more satisfactory than an essay test. Brownell points out many instances where subjectivity and judgment enter into the construction, administration, and use of objective tests. Objective tests are

objective in that competent individuals agree in the scoring. Objectivity is an important characteristic of evidence. However, it should not be purchased at the expense of the primary characteristic, validity. This leads us to another major problem in preparing examinations.

Practicability of Examinations

The ease with which evidence of the achievement of an objective of teaching can be collected and evaluated is an important consideration in an evaluation program. A test which is easily administered and quickly scored is more likely to be used than one in which the evidence is difficult to collect and time-consuming to score. There is much evidence to show that important outcomes of teaching are not evaluated because practicable methods are not available for collecting evidence of progress.

Practicability is so much desired that teachers have eagerly accepted, easily administered, and easily scored paper and pencil examinations and have assumed that they were valid. They have not asked about each test, Does it require of students the kinds of behavior which represent the objectives of my course, or does it give a satisfactory index of this behavior?

Index devices make an evaluation program more practicable. Although they do not yield direct evidence of desired behavior, they are valuable when they give a satisfactory index of the direct evidence. In college chemistry courses, an important objective is the ability to apply chemical principles. Hendricks, R. W. Tyler, and Frutchey (413) reported research in checking an index device with direct evidence of this outcome. An objective of French instruction is the ability to pronounce French words. Direct evidence of this ability may be collected about each student as he reads and pronounces French words. It is a time-consuming job to obtain this kind of evidence for each student and have it evaluated by two or more trained judges. Tharp (448) experimented with a paper and pencil index device which can be administered in class in five minutes and scored in less time. He made phonographic recordings of students' French pronunciations and had the records played and the pronunciations judged by three French teachers. The students also took the index examination. The correlation between the average ratings of the three French teachers and scores on the index examination was .84. If this relationship continues to hold from time to time with other groups of students and teachers, a practicable device is available for collecting evidence of the ability to pronounce French. As W. S. Monroe (434) pointed out it is misleading to assume that the relationship between direct and indirect tests is fixed. The assumption should be checked frequently.

Item Validity

Several methods of item validation have been used in selecting items for a test. Kelley and Krey (422) proposed that a good item for inclusion in a test is one whose "difficulty" decreases as the school grade increases.

Kelley pointed out that "items selected in this manner can be built into a final test which does consistently measure something, but whether it is what the test label calls for remains a matter of judgment."

Another method used in item validation is the degree to which an item discriminates between the better and poorer students. Differences have arisen in choosing an appropriate criterion for selecting the better and poorer students. In some cases it is assumed that the total score on the general achievement scale is the valid criterion. According to Lindquist and Anderson (427), "an item may be said to be perfect in discriminating power when every pupil who responds correctly to the item ranks higher on the general achievement scale than any pupil who fails on the item."

This method of statistical validation of test items has been used to pick out and eliminate from a group of test items those of low "validity" so that the test consisting of the remaining items will have higher "validity." Smith (442) proposed to test this hypothesis and concluded from his evidence that "so far as obtaining more and more valid instruments of measurement is concerned, statistical evaluation of individual items apparently has little to contribute."

The purpose of these methods of statistical validation is to construct a test which is consistent with respect to an internal criterion. When the criterion is the total score on the test, it is assumed that the test itself is valid. The question then arises, What makes the test itself valid? It appears that the answer to that question goes back to the meaning of the objective—the behavior expressing the objective, the situations in which the behavior may occur, and the evaluation of the behavior. The essential characteristic of validity is the kind of evidence which expresses the purpose of the evaluation.

Statistical methods are useful in making a test internally consistent or homogeneous. The process selects items which are highly related to a whole group. The kinds of items selected by the procedure are probably those kinds which are greatest in number in the total test and which have been given the most weight in the scoring. For example, in a test on educational statistics, including many items on computation and a few on interpretation, statistical validation of the items would discard the items on interpretation and tend to make the test internally consistent in respect to computation. Some of the computation items may be discarded also. Statistical validation in this sense destroys the value of the test for the purposes for which it was constructed. The process eliminated opportunities to get evidence of the ability of the students to interpret the use of statistical measures and some important kinds of behavior in computation.

Statistical methods are also used in selecting items by checking each item with an outside criterion. Horst (416) reported their use in developing a test for selecting salesmen. The external criterion was success on the job. The purpose was to build a test which ranked the employees in the same order as did the measure of success of one's work on the job.

From an original group of tests given the employees, items were selected which correlated highly with their success on the job and correlated but slightly with each other. In this way a new test was built which had a very high predictive value of success of the job. The validity of the new test rests upon the degree to which it is a good *index* of the valid kind of behavior—that which was used as evidence of success on the job.

Research Considerations

Lehman and Witty (426) gave seven major objectives of an introductory course in psychology and pointed out seven obstacles involved in formulating objectives. These difficulties are founded on certain assumptions which often are not necessary to make. For example, a difficulty is "the disparity among the objectives that have been published." The assumption underlying this obstacle is that *all* introductory courses in psychology should have the same objectives. Fortunately, it is not necessary to make this assumption in setting up objectives. Teachers or departments may formulate the purposes for their own courses. Another difficulty, "our limited knowledge of the psychology of learning," is based on the assumption that we must know how learning takes place before we can find evidences that it has taken place. This assumption also is not necessary. If we can detect evidences of changes, it is not necessary to know how the changes have occurred, for the purposes of formulating objectives. Research in improving examinations must consider the assumptions underlying an appraisal program.

The development of testing technics likewise involves a series of assumptions. R. W. Tyler (453) pointed out some of these assumptions and suggested procedures for checking them. "If an appraisal program is to develop a sound body of philosophy, techniques, data, inferences, and principles, it is essential that these assumptions be tested one by one, so that we may be able to separate findings and methods which are based upon valid assumptions from those which are untenable because the assumptions underlying them are not true. This is a task which challenges all who work in the field of achievement test construction."

BIBLIOGRAPHY ON EDUCATIONAL TESTS AND THEIR USES

Chapter I. Educational Tests and Measurements in China, England, France, and Germany

A. CHINA

1. AI, J. A. W. *A Preliminary Report on English Tests in Senior High Schools*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
2. CHOU, SIEGEN. *An Analysis of the Judgments of Americans as to the Position of Chinese Characters*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
3. CHOU, SIEGEN, and CHEN, H. P. *General and Specific Color Preferences of Chinese Students*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
4. CHOW, H. C. *A Comparative Study of Two Methods of Teaching Decimal Multiplication*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
5. CHOW, H. C. *A Critique of the T. B. C. F. System*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
6. HSIAO, HSIAO HUNG. *An Experimental Study of the Eye-Hand Coördination*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
7. HSIAO, HSIAO HUNG. *The First Chinese Revision of the Pressey X-O Tests*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
8. HSIAO, HSIAO HUNG. *A Genetic Study of Memory Patterns*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
9. HSIAO, HSIAO HUNG. *Studies in Motion*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
10. OU-NI-LIN, MME. "The Binet-Simon Intelligence Tests Revised in China." *Revue Scientifique du Travail* 2: 443-61; 1930.
11. SHEN, EUGENE. *An Attempt To Apply Tests of Learning Ability to the College Entrance Examination*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
12. TSANG, CHIU-SAM, and CHUANG, CHAI-HSUAN. "Educational Research and Experimentation in China." *International Education Review*, 1933-34, Part 3. p. 351-61.
13. TSAO, ZEN-CHEN, and SHEH KWAN CHUN. *Progress in Writing as a Function of Practice Distribution*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).
14. WOO, C. F. *Personal Tempo and Speed in Some Rate Tests*. Nanking, China: Society of Tests and Measurements (74 Hung Miao).

B. ENGLAND, FRANCE, and GERMANY

15. ABADI, E. "Intelligence and Personality." *Psychologie et la Vie* 5: 245-48; 1931.
16. ABRAHAMSON, J. "An Attempt To Standardize Two Tests of Imagination and Observation." *Journal de Psychologie Normale et Pathologique* 24: 375-79; 1927.
17. BALLARD, PHILIP BOSWOOD. *The New Examiner*. London: University of London Press, 1925.
18. BARTSCH, KARL. "Die psychischen Leistungen der Kinder einer vierten Volksschulklasse verglichen mit den Leistungen einer Zweiten Hilfsschulklasse. (A Comparison of Achievement in Two Types of Schools.)" *Zeitschrift für pädagogische Psychologie* 30: 194-98; April, 1929.
19. BLANCO, W. "Tests of Intelligence." *Annales d'Orientation Professionnelle* 4: 341-43; 1931.
20. BLUMENFELD, WALTER. "Über quantitative und qualitative Bewertung von Test Leistungen. Vortrag auf dem 7 Internationalen Psychotechnischen Kongress in Moskau. (Concerning the Quantitative and Qualitative Evaluation of Test Performance. A Report from the Seventh International Psychotechnical Congress in Moscow.)" *Zeitschrift für angewandte Psychologie* 40: 209-30.

21. BOBERTAG, OTTO. "Bermerkung zu Lietzmann's Mitteilung 'Über den prognostischen Wert von Test und Zensur.' (Note with Reference to Lietzmann's Communication Concerning the 'Prognostic Value of Tests and School Reports.')" *Zeitschrift für pädagogische Psychologie* 34: 478-79; December, 1933.
22. BOBERTAG, OTTO. "Ergebnisse einer Versetzungs Statistik in Preussen. (Results of a New Type of Statistics in Prussia.)" *International Education Review*, 1933-34, Part 1. p. 103-5.
23. BOBERTAG, OTTO. "Leistungsschätzung und Leistungsmessung in der Schule. Ein Beitrag zur Frage 'Was leistet unsere Volksschule?'" *Zeitschrift für pädagogische Psychologie* 34: 377-93; October, 1933.
24. BOBERTAG, OTTO. "Variabilität und Konstanz von Begabung und Schulleistung. Variability and Constancy of Talents and School Performance.)" *Zeitschrift für pädagogische Psychologie* 32: 12-27; January, 1931.
25. BONNIS, L. "The Development of Intelligence in Backward Children." *Hygiène Mentale* 23: 197-202; 1928.
26. BRADFORD, E. J. G. "The Measurement of Perspective in the Geographical Outlook of Secondary School Pupils." *British Journal of Educational Psychology* 2: 332-52; November, 1932.
27. BRANDICOURT, VIRGILE. "Une Dictée Muette." *Bulletin, Société A. Binet*, August and September, 1926. (Reviewed in *L'Education* 18: January, 1927.)
28. BRAUN, FRITZ. "Von Einfluss des Schulalters auf die Schulleistungen. (The Influence of School Age on School Performance.)" *Archiv für die gesamte Psychologie* 70: 1-12; June, 1929.
29. BÜHLER, CHARLOTTE, and HETZER, HILDEGARDE. *Kleinkindertests*. Leipzig: J. A. Barth. (For children from 1-6.)
30. BURT, CYRIL. *Test in Fundamentals of Arithmetic*. London: London County Council.
31. CATTELL, R. B. "On the Measurement of Perseveration." *British Journal of Educational Psychology* 5: 76-92; February, 1935.
32. CHAMPNEYS, MARY C., compiler. *An English Bibliography of Examinations, 1900-1932*. London: Macmillan Co., 1934. 166 p. (Reviewed in *British Journal of Educational Psychology* 4: 333; 1934.)
33. CLAPARÈDE, ÉDUARD. *Comment Diagnostiquer les Aptitudes chez les Écoliers. (How to Diagnose Aptitudes of Students.)* Paris: E. Flammarion, 1924. 300 p.
34. CLAPARÈDE, ÉDUARD. "What To Think of Mental Tests?" *Nouvelles Littéraires*. June 14, 1930.
35. CRUCHET, R. "The Measurement of Intelligence in the Child from Birth to 2 or 3 Years of Age." *Journal de Médecine de Bordeaux* 107: 951-60; 1930.
36. DERRAY, E. "Les Leçons de l'Étranger dans ledépistage et l'éducation des arrières et des anormaux. (Lessons from Foreign Countries in the Identification and the Education of Backward and Abnormal Children.)" *L'Education* 20: 426-33; April, 1929.
37. DECROLY, O. "Application du Test de Ballard dans les Ecoles Belges. (The Use of Ballard's Tests in the Schools of Belgium.)" *L'Education* 21: 493; May, 1930.
38. DECROLY, O. "An Attempt at the Application of Ballard's Test in the Belgian Schools." *L'Année Psychologique* 27: 57-93; 1926.
39. DECROLY, O., and SEGERS, J. E. *An Attempt To Apply the Ballard Test*. Documents Pédotechniques, Vol. 11, No. 1. Brussels: Société Belge de Pédotechnie, 1932. 149 p.
40. DECROLY, O., and BUYSE, RAYMOND. *La Pratique des Tests Mentaux. (The Use of Mental Tests.)* Paris: F. Alcan, 1928. 402 p.
41. DECROLY, O., and BUYSE, RAYMOND. *The Use of Mental Tests*. Brussels: Maurice Lamartin, 1924. 60 p. (Reviewed in *L'Education* 16: 1924-25.)
42. DECROLY, O. "The Value of the Intelligence Quotient in Abnormal Children." *Journal de Neurologie et de Psychiatrie* 30: 985-88; 1930.
43. DELVAUX, A. *Contrôle de la Stanford Revision de Terman avec une préface du O. Decroly. (Use of the Stanford Revision by Terman with Preface by Decroly.)* Brussels: Lamartin, 1932. 203 p. (Reviewed in *L'Education* 24: 566-67; June, 1933.)
44. DELVAUX, A. *A Test of Terman's Stanford Revision on Children of Different Social Environment*. Documents Pédotechniques, Vol. 11, No. 2. Brussels: Société Belge de Pédotechnie, 1932. 201 p.

45. DE MONTLEBERT, SIMONNE ROUX. *The Determination of Aptitudes by the Method of Tests*. Paris: Delachaux, 1926. 152 p. (Reviewed in *L'Education* 17: 1925-26.)
46. DILGER, J. "Gruppeneinteilung bei der Gausschen Kurve. (Classification by Means of the Gaussian Curve)." *Industrielle Psychotechnik* 7: 258-64; August-September, 1930. 8: 225-28; August, 1931.
47. DUTHIL, RENÉ. "L'adaptation des Programmes et des Méthodes d'enseignement à l'Intelligence des élèves. (The Adaptation of the Programs and Methods of Teaching to the Intelligence of Pupils.)" *L'Education* 19: 344-49; March, 1928.
48. DUTHIL, RENÉ. "Comment la méthode des Tests nous revient d'Amérique plus riche et plus pratique. (How the Method of Testing Comes Back to Us from America Enriched and More Practical.)" *Revue Pédagogique* 85: 402-17; December, 1924.
49. DUTHIL, RENÉ. "L'enseignement de l'orthographe et la méthode des Tests." (The Teaching of Orthography and the Methods of Testing.)" *Revue Pédagogique* 86: 421-44; June, 1925.
50. DUTHIL, RENÉ. "Group Test in Arithmetic." *L'Education* 18: 420-26; April, 1927.
51. DUTHIL, RENÉ. "Initiation à la Méthode des Tests. (Initiation into Testing Methods.)" *L'Education* 17: 89-97; October, 1925.
52. DUTHIL, RENÉ. "The Measurement of Intelligence." *Psychologie et la Vie* 1: 12-14; 1927.
53. DUTHIL, RENÉ. "Notes de Classe et de Composition. (Notes Covering Classes Studying Composition.)" *L'Education* 18: 546-49; June, 1927.
54. DUTHIL, RENÉ. "La Place Exacte de la Méthode des Tests dans l'Enseignement. (The Exact Place of the Testing Methods in Teaching.)" *L'Education* 18: 395-98; April, 1927.
55. DUTHIL, RENÉ. "Tests Collectif d'Intelligence. (Group Intelligence Tests.)" *L'Education* 18: 398-406; April, 1927.
56. DUTHIL, RENÉ. "Tests d'Instruction: Orthographe. (Tests of Instruction: Spelling.)" *L'Education* 18: 406-20; April, 1927.
57. DWELSHAUVERS, G. "The Sense of the Concrete and of Intelligence as a Whole, or Factor G." *Psychologie et la Vie* 2: 189-90; 1928.
58. EAGLESHAM, ERIC. "A Comparison of the Effects on Retention of Various Methods of Revision." *British Journal of Educational Psychology* 1: 204-14; June, 1931.
59. EDER, JOSEF. "Über Sorgfaltsdiagnose. (The Diagnosis of Carelessness.)" *Zeitschrift für pädagogische Psychologie* 33: 165-69; April-May, 1932.
60. EGGINK, H., and BRADFORD, E. J. G. "Discussion on Measuring Geographical Perspective." *British Journal of Educational Psychology* 3: 183-86; June, 1933.
61. EKENBERG, MADELINE. "Zur Frage der Character Tests und ihrer Methodik. (Character Tests and Their Technique.)" *Zeitschrift für angewandte Psychologie* 34: 494-511.
62. EL KOUSSY, A. H. "A Note on Greys Analogy Test." *British Journal of Educational Psychology* 4: 294-95; November, 1934.
63. EMERY, L. "The Social Value of Selection." *Pour l'Ere Nouvelle* 72: 278-79; 1931.
64. EMMETT, W. G. "The Tetrad Criterion and Scholastic Examinations." *British Journal of Educational Psychology* 5: 93-100; February, 1935.
65. FARMER, ERIC. "The Predictive Value of Examinations and Psychological Tests in the Skilled Trades." *British Journal of Educational Psychology* 4: 47-55; February, 1934.
66. FAUVILLE, A. "General Intelligence and Psychographic Examination." *Journal de Psychologie Normale et Pathologique* 26: 123-30; 1929.
67. FAY, H. M. "L'Enseignement et l'Education des Anormaux en France. (The Instruction and Education of Abnormals in France.)" *L'Education* 24: 91-105; November, 1932.
68. FESSARD, A., and PIÉRON, HENRI. "Concerning the Number of Measures Necessary for the Standardization of a Test for a Psychometric Purpose." *L'Année Psychologique* 31: 246-59; 1930.
69. FESSARD, A., and PIÉRON, HENRI. "The Idea of Validity." *L'Année Psychologique* 31: 217-28; 1930.
70. FESSARD, A. "The Interpretation of Numerical Results in Aptitude Examinations." *Bulletin. Institut National d'Orientation Professionnelle* 1: 229-39; 1929.
71. FESSARD, A. "The Precision and Coherence of Test Results." *L'Année Psychologique* 28: 205-35; 1927.

72. FESSARD, A. B., and FESSARD, A. "Musical Aptitude and the Seashore Tests." *Bulletin, Institut National d'Orientation Professionnelle* 3: 1-11, 29-41; 1931.
73. FISCHLER, E., and ULLERT, I. "A Contribution to the Study of Tests for Immediate Memory." *Archives de Psychologie* 21: 293-306; 1929.
74. FOUCAULT, MARCEL. *The Measurement of the Intelligence of Students*. Paris: Delagrave, 1933. 137 p.
75. FOUCAULT, MARCEL. "La Memoire des Mots chez les Ecoliers de 10 à 12 Ans. (The Retention of Words by Pupils of 10 to 12 Years of Age.)" *L'Education* 16: 65-74; November, 1924.
76. FOUCAULT, MARCEL. "Qualities du Travail Mental, Loi d'Exercice et de Fatigue. (The Quality of Mental Work, Law of Exercise, and Fatigue.)" *L'Année Psychologique*, 1926. (Quoted in *L'Education* 21: 492; May, 1930.)
77. FRICKX, H. *Second Contribution to the Study of Methods in Tests*. Documents Pédotechniques, Vol. 9, No. 1. Brussels: Société Belge de Pédotechnie, 1930. 27 p.
78. FRICKX, J. "The Simon P.V. Test, Second Part." *Bulletin Trimestrielle de l'Office Intercommercial d'Orientation Professionnelle de Bruxelles* 9: 22-34; 1930.
79. GAMS, M., and SALKIND, A. "A Contribution to the Study of Certain Attention Tests." *Archives de Psychologie* 21: 307-19; 1929.
80. HAGUENIN, ELIZABETH. "Une Visite à l'Ecole Decroly. (A Visit to the Decroly School.)" *L'Education* 21: 333-45; March, 1930.
81. HALLER, M. *Du Choix des Tests dans la Détermination Pratique de l'Age Mental. (The Choice of Tests in the Practical Determination of Mental Age.)* Paris: F. Alcan.
82. HENNING, H. "Character Tests. New Apparatus and Methods for Two Persons." *Hygiène Mentale* 24: 88-93; 1929.
83. "Hochschulpädagogische Leistungsprüfungen. (Tests of Performance in Schools of College Rank.)" *Zeitschrift für angewandte Psychologie* 38: 77-248, 448-509.
84. HUTH, ALBERT. "Ein Testheftverfahren mit Qualitativer Bewertung. (A Test Procedure with Qualitative Evaluation.)" *Zeitschrift für pädagogische psychologie* 35: 225-28; 1934.
85. IZARD, J. *The Determination of Intellectual Development in Children. Its Clinical Importance*. Thèse de Médecine de Paris, 1930. 120 p.
86. JEANJEAN, G. "Les Anormaux et les sous-normaux dans l'Enseignement secondaire. (The Abnormal and the Sub-Normal in the Secondary Field.)" *L'Education* 20: 415-26; April, 1929.
87. JUHÁSI, ANDOR. "Die 'Krise' der Psychotechnik. (The Crisis of Psychotechnique.)" *Zeitschrift für angewandte Psychologie* 33: 456-64.
88. KENNEDY, FLORA. "The Practical Value of the June Downey Will-Temperament Tests." *British Journal of Educational Psychology* 4: 260-63; November, 1934.
89. KERN, BENNO, and LINDOW, M. P. J. "Die Mathematische Auswertung empirisch gefundener Kurven mit besonderer Berücksichtigung der Übungskurven. (The Mathematical Evaluation of Empirically Established Curves with Special Consideration of Practice Curves.)" *Zeitschrift für angewandte Psychologie* 35: 497-529; 1930.
90. KESSELRING, MICHAEL. "Über den Ausbau von Aufnahmeprüfungen, besonders für Aufbauschulen. (The Improvement of Entrance Examinations Especially for Teacher-Training Institutions.)" *Zeitschrift für pädagogische Psychologie* 30: 31-42, 86-99; January and February, 1929.
91. KORN, G. *Rechnenleistung und Rechenfehler. (Number Performance and Error.)* Berlin: Institut für Angewandte Psychologie.
92. KOVARSKY, VERA. "L'Inspection Psychologique à Montpellier. (Psychological Inspection at Montpellier.)" *L'Education* 21: February, 1930.
93. KOVARSKY, VERA. *The Measurement of Mental Capacities in Children and Adults, Normal and Abnormal; the Method of the Psychological Profile*. Paris: F. Alcan, 1927. 178 p.
94. KOVARSKY, VERA. "The Role of the Psychological Profile in Orthopsychiatry." *Annales Medico-Psychologiques* 88: 142-48; 1930.
95. KOVARSKY, VERA. "The Role of the Psychological Profile Method on Psychological Orthopedics." *Psychologie et la Vie* 3: 172; 1929.
96. KOVARSKY, VERA. "Some Applications of the Use of the Psychological Profile." *Journal de Psychologie Normale et Pathologique* 27: 805-15; 1930.
97. LAHY, J. M. "A Test of Logical Intelligence." *Travail Humain* 1: 129-51; 1933.

98. LÄMMERMANN, HANS. "Bericht über die Eichung einer Serie von Gruppentests für Acht- bis Vierzehnjährige Volksschüler." *Zeitschrift für angewandte Psychologie* 27: 1-47; April, 1926.
99. LÄMMERMANN, HANS. "Die Konstanz und die Übbarkeit von Denkleistungen Übungsversuche mit Intelligenz Tests. (The Reliability of Intelligence Test Scores; and Experiment with Reference to Practice Effect.)" *Zeitschrift für angewandte Psychologie* 46: 1-87.
100. LATZARUS, MARIE-THERÈSE. "Comment determiner le niveau intellectuel d'un enfant. (How to Determine the Intellectual Level of an Infant.)" *L'Education* 22: 272-74; February, 1931.
101. "Leistungsmessung der Grundschule." *Zeitschrift für pädagogische Psychologie* 34: 44-45; January, 1933. (A review of tests by OTTO BOBERTAG.)
102. LEOPOLD, KATHLEEN B. "The Effect of Creative Work on Aesthetic Appreciation: An Experiment in Teaching Poetry." *British Journal of Educational Psychology* 3: 42-64; February, 1933.
103. LEVITOF, N. "Some Factors of Variation in Test Results." *Revue Scientifique du Travail* 2: 197-218; 1930.
104. LIEFMANN, ELSE. "Volksschülerinnen ihre Geistigen und Körperlichen Leistungen und die Beziehung zur Konstitution." *Zeitschrift für angewandte Psychologie* 42: 102-216.
105. LIETZMANN, W. "Über den prognostischen Wert von Test und Zensur. (The Prognostic Value of Tests and School Reports.)" *Zeitschrift für pädagogische Psychologie* 34: 296-99; July-August, 1933.
106. LIETZMANN, W. "Von der Amerikanischen Testbewegung. (The American Test Movement.)" *Zeitschrift für pädagogische Psychologie* 32: 476-86; November, 1931.
107. LINDBECK, CARL. "Untersuchungen über die Arbeitskurve von Schulkindern bei fortlaufender schriftlicher Rechenarbeit. (Investigations Concerning the Work Curve of School Children, with Continued Written Number Work.)" *Zeitschrift für pädagogische Psychologie* 34: 223-33; June, 1933.
108. LIPMANN, OTTO. "Der Oskulationswert. (The Osculation Value.)" *Zeitschrift für angewandte Psychologie* 45: 131-39.
109. LOOSLI-USTERI, M. "Application of the Rorschach Test to Different Groups of Children from 10 to 13 Years of Age." *Archives de Psychologie* 22: 51-106; 1929.
110. LOOSLI-USTERI, M. "Interpretations of the Rorschach Test." *Archives de Psychologie* 23: 349-65; 1932.
111. MAINWARING, JAMES. "Tests of Musical Ability." *British Journal of Educational Psychology* 1: 313-21; November, 1931.
112. MASSON, OURSEL P. "The Use of Tests for Mental Exercise." *Proceedings of the Fifty-Third Session of the Association Française pour l'Avancement des Sciences*. Paris: the Association, 1929. p. 634-35.
113. "The Measurement of the Child's Intelligence." *Annales de l'Enfance* 4: 869-81; 1930.
114. MEILL, R. "Concerning the Factor Theory." *Archives de Psychologie* 22: 313-27; 1930.
115. MEILL, R. "Studies of the Forms of Intelligence." *Archives de Psychologie* 22: 201-81; 1930.
116. MÉNESSIER, P. "De l'emploi des tests en Orientation Professionnelle. (The Use of Tests in Professional Orientation.)" *L'Education* 22: 604-5; July, 1931.
117. MENNENS, G. "Experimental Study of Different Mental Aptitudes in Prisoners." *Journal de Psychologie* 28: 283-302; 1931.
118. "La Méthode Decroly. (The Decroly Method.)" *L'Education* 21: 151-60; December, 1929.
119. MIRA, E. "A New Test for the Exploration of Affectivity." *Annales d'Orientation Professionnelle* 4: 357-59; 1931.
120. MOERS, MARTHA. "Die Logische Bedeutung der 'Gleich Möglichen Fälle' im Nenner des Wahrscheinlichkeitsbruches. (The Logical Meaning of 'Equally Probable Case' in the Denominator of a Fraction Denoting Probability.)" *Archiv für die gesamte Psychologie* 77: 249-64; 1930.
121. MOERS, MARTHA. "Zur Prüfung des Sittlichen Verständnisses Jugendlicher. (Testing Moral Understanding among Youth)." *Zeitschrift für angewandte Psychologie* 34: 431-60; 37: 56-73; 1930.

122. MONCHAMPS, MILE., and MORITZ, MILE. *Les Etapes Mentales de l'Observation des Images. (Mental Stages in the Observation of Pictures.)* Brussels: Oeuvre Nationale de l'Enfance (67 Avenue de la Toison-d'Or). 178 p. (Reviewed by R. DUTHIL in *L'Education* 19: 369-71; March, 1928.)
123. MONTASSUT, N. M. "Le Docteur Decroly—Sa Personnalité, ses Travaux, son Oeuvre. (Dr. Decroly—His Personality, Labors, and Work.)" *L'Education* 21: 129-36; December, 1929.
124. MÜLLER, ANDREAS. "Abhängigkeit der Schulleistungen von wirtschaftlichen und sozialen Einflüssen. (Relation of School Performance to Business and Social Influences.)" *Archiv für die gesamte Psychologie* 83: 119-96; January, 1932.
125. MÜLLER, CARL VIKTOR. "Experimentelle Untersuchungen über kindliche Schlussprozesse bei Schlüssen nach der 4 Figur. (Experimental Investigations of Child Reasoning with Syllogisms of the Fourth Type.)" *Archiv für die gesamte Psychologie* 86: 407-58; September, 1932.
126. MÜLLER, CARL VIKTOR. "Experimentelle Untersuchungen über kindliche Schlussprozesse mit Besonderer Berücksichtigung der Vorgänge der Repräsentation. (Experimental Investigations Concerning the Reasoning of Children with Special Reference to Type of Representation.)" *Archiv für die gesamte Psychologie* 78: 379-494; 79: 1-166; 1930.
127. MYERS, CHARLES S. "Psychological Cautions in the Use of Statistics." *Zeitschrift für angewandte Psychologie* 36: 82-86; 1930.
128. NASSRI, K. "Application of the Pintner-Paterson Test to the Pupils of Five Certificate Classes." *Encephale* 27: 232-41; 1932.
129. NASSRI, K. *Intelligence Tests and Scholastic Achievement*. Paris: Presses Universitaires, 1930. 246 p.
130. NIHARD, R. *The Method of Tests*. Juvesy, Seine et Oise: Editions du Cerf, 1932. 240 p.
131. NORDEN, IRMGARD. *Binetarium*. Berlin: Albrecht Dürer.
132. NORDEN, IRMGARD. "Eine Neuberbeitung der Binet-Methode. (A New Revision of the Binet Method.)" *Zeitschrift für Kinderforschung* 37: 75-92.
133. PARSCHAU, FRANZ. "Prüfung des Dauerfolges der Schularbeit. (Testing the Permanence of School Work.)" *Zeitschrift für pädagogische Psychologie* 29: 351-56; July-August, 1928.
134. PAULI, R., and WENZL, A. "Ein Einfachstes Verfahren zur Berechnung Korrelation Zusammenhänge. (A Very Simple Method for the Calculation of Coefficients of Correlation.)" *Zeitschrift für pädagogische Psychologie* 35: 113-20; 1934.
135. PEDERSEN, M. "The Dearborn Test Applied to Norwegian Pupils." *Archives de Psychologie* 23: 179-88; 1931.
136. PEYRAUBE, J. "Les Méthodes allemandes de Selection scolaire. (The German Methods of School Selection.)" *Revue Pédagogique* 87: 97-117; August, 1925.
137. PIÉRON, HENRI. *Le Développement Mental et l'Intelligence. (Mental Development and Intelligence.)* Paris: F. Alcan, 1929. 95 p.
138. PIÉRON, HENRI. "Experimental Criticism of Examination Methods." *Bulletin, Société Française de Pédagogie* 27: 20-26; 1928.
139. PIÉRON, HENRI, and PIÉRON, MME. H. "Making and Utilizing an Intelligence Test in Professional Orientation." *Hygiène Mentale* 23: 1-3; 1928.
140. PIÉRON, HENRI. "The Problem of Intelligence." *Scientia* 41: 337-48; 1927.
141. PIÉRON, HENRI. "Scales of Development and Evaluation of Intelligence; Necessity of Mental Profile in Professional Orientation." *Annales de l'Enfance* 1: 254-61; 1927.
142. PIÉRON, HENRI. "Theoretical and Practical Aspects of the Problem of Intelligence." *Kwartalnik Psychologiczny* 3: 1-8; 1932.
143. PIÉRON, MME. H. "A French Standardization of the Barcelona Test." *L'Année Psychologique* 29: 113-41; 1928.
144. PIÉRON, MME. H. "An Intelligence Test for Professional Orientation; Its Standardization." *L'Année Psychologique* 27: 174-202; 1926.
145. PIÉRON, MME. H. "Psychotechnical Researches in the School and the Collaboration of Educators." *Pour l'Ere Nouvelle* 73: 279-82; 1926.
146. PIÉRON, MME. H. "Standardization of a Test of Attention. (Toulouse and Piéron's Test.)" *Bulletin, Institut National d'Orientation Professionnelle* 1: 105-12; 1929.
147. PIÉRON, MME. H. "Standardization of Memory Tests from the Psychological Records of Professional Orientation." *Bulletin, Institut National d'Orientation Professionnelle* 2: 60-64; 1930.

148. PIÉRON, MME. H. "The Standardization of Tests: Numerical Series Tests." *Bulletin, Institut National d'Orientation Professionnelle* 1: 40-44, 61-64, 87-94; 1929.
149. PIÉRON, MME. H. "A Standardization of Two Memory Tests from the Psychological List." *Bulletin, Institut National d'Orientation Professionnelle* 2: 8-14; 1930.
150. PIÉRON, MME. H. "Une Test d'Intelligence pour l'Orientation Professionnelle." *L'Education* 21: 493; May, 1930.
151. POPPELREUTER, WALTHER. "Die methodische Rolle der 'determinierenden Tendenzen' bei Begutachtungsexperimenten. (The Methodical Role of Deterministic Tendencies in Experiments Involving Personal Opinion.)" *Archiv für gesamte Psychologie* 83: 385-95; February, 1932.
152. *Le Problème de l'Intelligence. (The Problem of Intelligence.)* Bulletin de la Société Française de Pédagogie, September, 1927.
153. RADECKE, W. "A Test of Intelligence for Adults." *Journal de Psychologie Normale et Pathologique* 24: 831-50; 1927.
154. RANGACHAR, C. "Differences in Perseveration among Jewish and English Boys." *British Journal of Educational Psychology* 2: 199-211; June, 1932.
155. REMY, MME. "Feuille d'Examen pour les Écolés Maternelles. Les Cours Préparatoires et Élémentaires." *Bulletin, Société A. Binet*, February-March 1926. (Reviewed in *L'Education* 19: 28-29; 1927.)
156. REVAULT D'ALLONES, G. "A Guide for the Psychological Examination of Normal People." *Psychologie et la Vie* 3: 11-13; 1929.
157. RÉVÉSZ, GÉZA. "Prüfung der Rechnerischen Fähigkeit und Fertigkeit an Schülern der Höchsten Klasse der Grundschule, I-II." *Zeitschrift für angewandte Psychologie* 36: 104-34; 215-36; 1930.
158. REYNIER, M. "Character Tests." *Pour l'Ere Nouvelle* 8: 117; 1929.
159. ROBSON, WINIFRED F. "The Vocabulary Burden in the First Year of French." *British Journal of Educational Psychology* 4: 264-93; November, 1934.
160. ROHRSCHACH, HERMANN. *Psychodiagnostik, Methodik und Ergebnisse eines Wahrnehmungsdiagnostischen Experiments. (Psycho-Diagnostic Methods and the Results of Experiments in Perception.)* Leipzig: 1921.
161. ROSENBLUM, A. "Kontrolltest als Prüfungsmethode der Testhomogenität. (A Control Test as a Measure of Test Homogeneity.)" *Zeitschrift für angewandte Psychologie* 40: 493-502; 1931.
162. SAINTE-LAGUE. "What Is a Test?" *Psychologie et la Vie* 2: 72-73; 1928.
163. SALUSCHNY, A. S. "Die Organisiertheit der Schulgruppen. (Social Participation among School Groups.)" *Zeitschrift für angewandte Psychologie* 33: 443-55; 1929.
164. SANDON, FRANK. "The Basis of Marking." *British Journal of Educational Psychology* 1: 296-312; November, 1931.
165. SANDON, FRANK. "The Necessary Imperfections of an Examination." *British Journal of Educational Psychology* 5: 180-93; June, 1935.
166. SANDON, FRANK. "Progress through a Secondary School as Measured by School Marks." *British Journal of Educational Psychology* 3: 269-90; November, 1933.
167. SANDON, FRANK. "Review of a Secondary School Test by W. A. Brockington (Oxford University Press, 1934. 64 p.)." *British Journal of Educational Psychology* 5: 106-9; 1935.
168. SCHLOTTE, FELIX. "Testhefter für die Auslese der Minderbegabten. (Test Sheets for the Selection of the Less Talented.)" *Pädagogisch Psychologische Arbeiten* 19: 1931. (Leipzig.)
169. SCHMIDBERGER, GUSTAVE. "Über Geschlechts Unterschiede in der Rechnen begabung. (Six Differences in Arithmetical Abilities.)" *Zeitschrift für pädagogische Psychologie* 33: 70-85, 140-65; 1932.
170. SEEMANN, JOHANN. "Untersuchungen über die Psychologie des Rechnens und der Rechenfehler. (Investigation Concerning the Psychology of Number and Number Errors.)" *Archiv für die gesamte Psychologie* 69: 1-180; May, 1929.
171. SEGERS, J. E. "Application of the B-D Group Tests." *Journal de Neurologie et Psychiatrie* 31: 168-74; 1931.
172. SEREBROWSKOJA, M. "Morphologischer Typus. (The Morphological Type.)" *Zeitschrift für Konstitutionslehre* 14: 1929.

173. SEREBROWSKOJA, M. "Über Körperliche Leistungsfähigkeit der Schulkinder Moskaus. (Concerning the Physical Performance of the Children of Moscow.)" *Zeitschrift für Schulgesundheitspflege und Soziale Hygiene* 42: 1929.
174. SIMON, TH. "Culture Test." *Bulletin, Société A. Binet* 21: 1-6; 1930.
175. SIMON, TH. *Test Collectif de Dessins Raisonnés. (Group Test in Planned Drawing.)* Paris: Hachette. (Reviewed by R. DUTHIL in *L'Education* 20: 449-50; April, 1929.)
176. SIMON, TH. "The Test of Battery A, or the Difficulties of Experimentation. *Bulletin, Société A. Binet* 32: 159-64; 1932.
177. SPEARMAN, CHARLES E. "Analysis of Abilities into Factors by the Method of Least Squares." *British Journal of Educational Psychology* 4: 183-85; June, 1934.
178. SPEARMAN, CHARLES E. "The Factor Theory." *Archives de Psychologie* 22: 313-27; 1930.
179. SPEARMAN, CHARLES E. "The Theory of Two Factors and That of 'Sampling.'" *British Journal of Educational Psychology* 1: 140-61; June, 1931.
180. "Standardization of the Stenquist Mechanical Aptitude Tests." *Bulletin, Institut National d'Orientation Professionnelle* 3: 121-36; 1931.
181. STEPHENSON, W. "An Introduction to So-Called Motor Perseveration Tests." *British Journal of Educational Psychology* 4: 186-208; June, 1934.
182. STERN, KÄTHE. *Experimentelle Untersuchung des Leseprozesses bei Anfängern. (Experimental Investigations of the Reading Process with Beginners.)* Breslau.
183. STERN, W. *Differentielle Psychologie.* Leipzig: 1921.
184. "Summary of Testing Methods in the Ecolé de Formation, Société de Strasbourg." *L'Education* 18: 426-27; 1927.
185. SZONDI, L. "Zur Psychometrie der Tests; Versuch einer kritischen Darlegung der Massbegriffe und Massinheiten der Tests. (Psychological Tests—An Experiment Concerning the Critical Exposition of Mass Concepts and Mass Unity of Tests.)" *Archiv für die gesamte Psychologie* 72: 43-114; October 1929.
186. *Test Collectif de Développement Mental. (Group Tests of Mental Growth.)* Nancy: Société d'Impressions Typographiques, 1926. (Reviewed in *L'Education* 18: 239; January, 1927.)
187. "Testserie zur Psychologischen untersuchung der Schulanfänger. (Test Series for the Psychological Investigation of Beginners in School.)" *Zeitschrift für pädagogische Psychologie* 35: 178-79; 1934. (Tests published by the Pädagogisch-psychologische Institut, Leipzig.)
188. THEISS, HERBERT. "Experimentelle Untersuchungen über die Erfassung des Handschriftlichen Ausdrucks durch Laien. (Experimental Investigations of Handwriting as an Index of Character.)" *Psychologische Forschung* 15: 276-358; 1931.
189. THOMSON, EDITH I. M. "A Study of the Efficiency of Individual Work." *British Journal of Educational Psychology* 2: 212-20; 257-75; June and November, 1932.
190. THOMSON, G. H. "Group Factors in School Subjects." *British Journal of Educational Psychology* 5: 194-99; June, 1935.
191. THOMSON, G. H. "The Standardization of Group Tests and the Scatter of Intelligence Quotients." *British Journal of Educational Psychology* 2: 92-112, 125-38; February and June, 1932.
192. THORNER, HANS. "Experimentelle Untersuchungen zur Psychologie des Lesens. (Experimental Investigations in the Psychology of Reading.)" *Archiv für die gesamte Psychologie* 71: 127-64, 165-84; August, 1929.
193. THOULESS, R. H. "A Fallacious Argument in Educational Psychology." *British Journal of Educational Psychology* 2: 196-98; June, 1932.
194. THUMB, NORBERT. "Der Faktorenaufbau einer Testreihe. (Factorial Structure of a Test Series.)" *Zeitschrift für angewandte Psychologie* 45: 86-130.
195. URBAN, F. M. "Die Methode des durchschnittlichen Fehlers. (The Method of Mean Error.)" *Archiv für die gesamte Psychologie* 74: 141-62; January, 1930.
196. URBAN, F. M. "Über die Methode der gleichen Abstufungen. (The Method of Equal Intervals.)" *Archiv für die gesamte Psychologie* 80: 291-312; May, 1931.
197. VALENTINE, C. W., and EMMETT, W. G. *The Reliability of Examinations.* London: University of London Press, 1932. 196 p. (Reviewed by T. P. NUNN in *British Journal of Educational Psychology* 2: 353-57; 1932.)

198. WALLIS, B. C. *The Measurement of Ability in Children*. New York: Oxford University Press, 1931. 36 p. (Reviewed in *British Journal of Educational Psychology* 2: 14.)
199. WEIGL, EGON. "Die Bedeutung der Testprüfung hinsichtlich der Frage des Übergangs von Grundschule zur Höheren Schule." *Zeitschrift für angewandte Psychologie* 33: 492-98.
200. WELLES, L. "The Selection of the Elite." *Psychologie et la Vie* 2: 174-76; 1923.
201. WILLIAMS, G. P. "The North Hampton Composition Scale (London: G. G. Harap and Co.)." *British Journal of Educational Psychology* 4: 121; 1934.
202. WILSON, J. H. "Group Factors among Abilities Involved in a School Certificate Examination." *British Journal of Educational Psychology* 3: 71-86, 99-108; February and June, 1933.
203. WILSON, J. H. "The Tetrad Criterion and Scholastic Examinations." *British Journal of Educational Psychology* 5: 101-5; February, 1935.
204. WINKLER. *Test serie zur Psychologischen Untersuchung der Schulneulinge. (Tests for the Psychological Investigation for Beginners in School.)* Leipzig: Pädagogisch-psychologischen Institut. (Reviewed in *Zeitschrift für angewandte Psychologie* 27: 395.)

Chapter II. Present Tendencies in the Uses of Educational Measurements

205. ALLEN, RICHARD D. "The Program of Measurement in the Secondary Schools of Providence." *Junior-Senior High School Clearing House* 8: 326-29; February, 1934.
206. AMERICAN COUNCIL ON EDUCATION. *Cumulative Record Card*. Washington, D. C.: the Council. (Reproduced in *Baltimore Bulletin of Education* 13: facing p. 16; September, 1934.)
207. AYER, FRED C. "Marks Have Value." *Journal of Education* 118: 157; March 18, 1935.
208. BAKER, FLORENCE, and BROOM, M. E. "Concerning One Criterion for the Choice of Primary Reading Tests." *Journal of Applied Psychology* 16: 419-20; August, 1932.
209. BRUECKNER, LEO J., and MELBY, ERNEST O. *Diagnostic and Remedial Teaching*. Boston: Houghton Mifflin Co., 1931. 598 p.
210. BURTON, WILLIAM H. *Supervision and the Improvement of Teaching*. New York: D. Appleton and Co., 1922. 510 p.
211. CARPENTER, W. W., and RUFI, JOHN. *The Teacher and Secondary School Administration*. Boston: Ginn and Co., 1931. 460 p.
212. CARTER, HOMER L. J. "An Attempt to Increase Reading Efficiency." *Educational News Bulletin* 2: 13-19; June, 1932. Kalamazoo, Mich.: Western State Teachers College.
213. CARTER, HOMER L. J. "Clinical Service at Western State Teachers College." *Papers of the Michigan Academy of Science, Arts, and Letters* 20: 539-47; 1934.
214. CHARLES, CECIL M. "Comparison of the Intelligence Quotients of Three Different Mental Tests Applied to a Group of Incarcerated Delinquent Boys." *Journal of Applied Psychology* 17: 581-84; October, 1933.
215. CHASE, HARRY WOBURN. "Administrator and the Testing Program." *Educational Record* 15: 29-39; January, 1934.
216. DOUGLASS, HARL R. "The Effects of State and National Testing on the Secondary School." *School Review* 42: 497-509; September 1934.
217. ENGLE, T. L. "Personality Study of a Group of High School Honor Society Pupils." *Journal of Applied Psychology* 18: 293-96; April, 1934.
218. EVANS, CALVIN O. "A Guidance Experiment in the Pratt, Kansas, High School." *American School Board Journal* 89: 18-65; August, 1934.
219. FINCH, FRANK H., and NEMZEK, C. L. "Prediction of College Achievement from Data Collected during the Secondary School Period." *Journal of Applied Psychology* 18: 454-60; June, 1934.

220. FRITZ, RALPH A. "Predicting College Marks and Teaching Success for Students in a Teachers College." *Journal of Applied Psychology* 17: 439-46; August, 1933.
221. GARRISON, NOBLE LEE. *The Technique and Administration of Teaching*. New York: American Book Co., 1933. 593 p.
222. GATES, ARTHUR I. *The Improvement of Reading*. New York: Macmillan Co., 1927. 440 p.
223. GILLINGHAM, ANNA. "What Is Measurable?" *Child Study* 12: 35-37; November, 1934.
224. GORDON, EDGAR B. "Teaching Music to School and Community by Radio." *Junior-Senior High School Clearing House* 7: 72-74; October, 1932.
225. HECK, ARCH O. "The Development of Various Pupil-Personnel Services." *Educational Research Bulletin (Ohio State University)* 14: 98-102; April 17, 1935.
226. HENRY, T. S. "William—A Behavior Problem." *Papers of the Michigan Academy of Science, Arts, and Letters* 20: 559-66; 1934.
227. HORN, ERNEST. "Another Chapter on Tests for the Volume of Conclusions and Recommendations." *Social Studies* 26: 13-22; January, 1935.
228. JENKINS, CLIFFORD D. *The Standard Graduation Examination for Elementary Schools as a Means of Predicting Success of Pupils in Certain High School Subjects*. Master's thesis, Pennsylvania State College, 1934. (Abstract in *Penn State Studies in Education*, No. 12. p. 27-28.)
229. JOHNSTON, J. B., chairman. "The 1933 College Sophomore Testing Program." *Educational Record* 14: 522-71; October, 1933. (Also printed separately.)
230. KELLEY, TRUMAN L., and KREY, A. C. *Tests and Measurements in the Social Sciences*. Report of the Commission on the Social Studies of the American Historical Association, Part IV. New York: Charles Scribner's Sons, 1934. 635 p.
231. LANDIS, CARNEY, and KATZ, S. E. "The Validity of Certain Questions Which Purport to Measure Neurotic Tendencies." *Journal of Applied Psychology* 18: 343-54; June, 1934.
232. LINCOLN, EDWARD A., and WORKMAN, L. L. *Testing and the Uses of Test Results*. New York: Macmillan Co., 1935. 317 p.
233. LONGSTREET, R. J. "An Experiment with the Thurstone Attitude Scales." *School Review* 43: 202-8; March, 1935.
234. MCCONN, MAX. "Educational Guidance: Progress toward Scientific Procedures." *School and Society* 40: 537-42; October 27, 1934.
235. MCCONN, MAX. "Educational Guidance Is Now Possible." *Educational Record* 14: 475-99; October, 1933.
236. MCCONN, MAX. "Measurement in Educational Experimentation." *Educational Record* 15: 106-19; January, 1934.
237. MACOMBER, F. G. "Marking System Rates an 'E'." *Journal of Education* 118: 35-37; January 21, 1935.
238. MAXWELL, C. R. "The Workbook; a Recent Development." *American School Board Journal* 88: 16, 44, 46; March, 1934.
239. MONROE, MARION. *Children Who Cannot Read*. Chicago: University of Chicago Press, 1932. 205 p.
240. MONROE, WALTER S., and STREITZ, RUTH. *Directing Learning in the Elementary School*. New York: Doubleday, Doran and Co., 1932. 480 p.
241. MOORE, HERBERT C., and STEELE, ISABEL. "Personality Tests." *Journal of Abnormal and Social Psychology* 29: 45-52; April, 1934.
242. NATIONAL EDUCATION ASSOCIATION, DEPARTMENT OF CLASSROOM TEACHERS. *The Classroom Teacher and Character Education*. Seventh Yearbook. Washington, D. C.: the Association, June, 1932. Chapter 6, "The Teacher and Individual Pupil Guidance," p. 135-68.
243. NATIONAL EDUCATION ASSOCIATION, RESEARCH DIVISION. "Education for Character, Part II: Improving the School Program." *Research Bulletin* 12: 81-144; May, 1934.
244. NEWELL, HORATIO W. "The Methods of Child Guidance Adapted to a Public School Program." *Mental Hygiene* 18: 362-72; July, 1934.
245. OLANDER, HERBERT T. "The Need for Diagnostic Testing." *Elementary School Journal* 33: 736-45; June, 1933.
246. O'SHEA, WILLIAM A. "An Improved Method of Administration in the Elementary School." *Education* 54: 362-67; February, 1934.

247. RAMSAY, CALVIN H. "Diagnostic Testing and Remedial Teaching in the Junior-Senior High School." *Junior-Senior High School Clearing House* 9: 232-34; December, 1934.
248. ROBB, EUGENE K. *A Study of the Results of the Direct and Incidental Methods of Instruction in the Field of Character Education*. Doctor's thesis, Pennsylvania State College, 1934. (Abstract in *Penn State Studies in Education* No. 12, p. 43-44.)
249. ROBINSON, BRUCE B. "The Place of the Child Guidance Clinic in Mental Hygiene." *Educational Method* 14: 180-83; January, 1935.
250. SAUVAIN, WALTER HOWARD. *A Study of the Opinions of Certain Professional and Non-Professional Groups Regarding Homogeneous or Ability Grouping*. Contributions to Education, No. 596. New York: Teachers College, Columbia University, 1934. 151 p. (Abstract in *Teachers College Record* 36: 145-46; November, 1934.)
251. SEACOE, MAY V. "The Evaluation of Certain Intelligence Tests." *Journal of Applied Psychology* 18: 432-36; June, 1934.
252. SEGEL, DAVID, and GERBERICH, J. R. "Differential College Achievement Predicted by the American Council Psychological Examination." *Journal of Applied Psychology* 17: 637-45; December, 1933.
253. SEGEL, DAVID. "Iowa's Testing Program." *School Life* 19: 49; November, 1933.
254. SEGEL, DAVID. "Measurement Today." *School Life* 20: 188-89; April, 1935.
255. SEGEL, DAVID. *National and State Cooperative High-School Testing Programs*. U. S. Department of the Interior, Office of Education, Bulletin, 1933, No. 9. Washington, D. C.: Government Printing Office, 1933. 47 p.
256. SEGEL, DAVID. "The Use of Tests and Measurements in the Evaluation of Instruction by Radio." *Junior-Senior High School Clearing House* 7: 75-77; October, 1932.
257. SMELTZER, C. H. "Educational Engineering in Testing and Diagnosis." *Educational Method* 12: 526-30; June, 1933.
258. STAGNER, ROSS. "The Intercorrelation of Some Standardized Personality Tests." *Journal of Applied Psychology* 16: 453-64; October, 1932.
259. STEPHENS, WINSTON B. "Tests and Student Guidance." *Junior-Senior High School Clearing House* 8: 341-46; February, 1934.
260. TERRY, PAUL W. "Prognostic Value of Different Types of Tests in Courses in Educational Psychology." *Journal of Applied Psychology* 18: 231-40; April, 1934.
261. TORGERSON, T. L., and AAMODT, GENEVA P. "The Validity of Certain Prognostic Tests in Predicting Algebraic Ability." *Journal of Experimental Education* 1: 277-79; March, 1933.
262. TROXELL, ELEANOR. "A Study of the Informational Background of Kindergarten Children as It Affects Reading Readiness." *Educational News Bulletin* 4: 2-6; May, 1934. Kalamazoo, Mich.: Western State Teachers College.
263. "The Tulsa Experiment with Ungraded Schools." *Elementary School Journal* 33: 733-34; June, 1933.
264. TYLER, RALPH W. "Evaluation: A Challenge and an Opportunity to Progressive Education." *Educational Record* 16: 122-31; January, 1935.
265. VAN LIEW, C. C. "Can the Workbook be Justified?" *School Executives Magazine* 53: 38-39; October, 1933.
266. WHITMER, CARROLL A. "Study of the Scholastic Progress of College Probationers." *Journal of Applied Psychology* 17: 39-48; February, 1933.
267. WILLIAMSON, E. G. "The Cooperative Guidance Movement." *School Review* 43: 273-80; April, 1935.
268. WOOD, BEN D. "Basic Considerations." *Review of Educational Research* 3: 5-20, 62; February, 1933.
269. WOOD, BEN D. "Coordinated Examining and Testing Programs." *Educational Record* 15: 48-55; January, 1934.
270. WOODY, CLIFFORD, and OTHERS. "A Symposium on the Effects of Measurement on Instruction." *Journal of Educational Research* 28: 481-527; March, 1935.
271. WRIGHTSTONE, J. WAYNE. "Correlations Among Tests of High-School Subjects." *School Review* 43: 198-201; March, 1935.

Chapter III. Objective Achievement Test Construction

272. ANASTASI, ANNE. "The Influence of Practice upon Test Reliability." *Journal of Educational Psychology* 25: 321-35; May, 1934.
273. ANDRUSS, HARVEY A. "True and False Correction Test." *Balance Sheet* 16: 61; October, 1934.
274. ARNOLD, J. N. "Monogram for Determining Validity of Test Items." *Journal of Educational Psychology* 26: 151-53; February, 1935.
275. BAWDEN, H. T. "An Economy Device for the Testing Program." *Educational Method* 13: 150-54; December, 1933.
276. BRIGGS, THOMAS H., and ARMACOST, GEORGE H. "Results of an Oral True-False Test." *Journal of Educational Research* 26: 595-96; April, 1933.
277. BROOKS, JOHN D. "A New Form of Objective Test." *Journal of Education* 115: 282; April 4, 1932.
278. BROWN, CLARENCE W; BARTELME, PHYLLIS; and COX, GERTRUDE M. "The Scoring of Individual Performance on Tests Scaled According to the Theory of Absolute Scaling." *Journal of Educational Psychology* 24: 654-62; December, 1933.
279. BROWNELL, WM. A. "On the Accuracy with Which Reliability May Be Measured by Correlating Test Halves." *Journal of Experimental Education* 1: 204-15; March, 1933.
280. BRUECKNER, LEO J., and HAWKINSON, MABEL J. "The Optimum Order of Arrangement of Items in a Diagnostic Test." *Elementary School Journal* 34: 351-57; January, 1934.
281. BRUECKNER, LEO J., and ELWELL, MARY. "Reliability of Diagnosis of Error in Multiplication of Fractions." *Journal of Educational Research* 26: 175-85; November, 1932.
282. CALDWELL, FLOYD F. "Speed as a Factor with Children of Superior and Inferior Intelligence." *Journal of Educational Research* 26: 522-24; March, 1933.
283. CAPRON, VIRGINIA LEE. "The Relative Effect of Three Orders of Arrangement of Items upon Pupils' Scores in Certain Arithmetic and Spelling Tests." *Journal of Educational Psychology* 24: 687-95; December, 1933.
284. CHICAGO UNIVERSITY. BOARD OF EXAMINATIONS. *Manual of Examination Methods*. Prelim. ed. Chicago: University of Chicago Bookstore, June, 1933. 177 p.
285. CLASS, E. C. "The Effect of the Kind of Test Announcement on Students' Preparation." *Journal of Educational Research* 28: 358-61; January, 1935.
286. COOK, WALTER WELLMAN. *The Measurement of General Spelling Ability Involving Controlled Comparisons Between Techniques*. University of Iowa Studies in Education, Vol. 6, No. 6. Iowa City: the University, 1932. 112 p.
287. COPELAND, HERMAN A. "A Note on the Effect of Teaching on the Reliability Coefficient of an Achievement Test." *Journal of Applied Psychology* 18: 711-16; October, 1934.
288. CRAWFORD, C. C. "How to Study for Objective Tests." *Education* 53: 413-16; March, 1933.
289. CUFF, NOEL B. "A New Device That Scores Tests." *Journal of Educational Psychology* 26: 73-77; January, 1935.
290. CUFF, NOEL B. "Scoring Objective Tests." *Journal of Educational Psychology* 23: 681-86; December, 1932.
291. CURETON, EDWARD E. "Validation against a Fallible Criterion." *Journal of Experimental Education* 1: 258-63; March, 1933.
292. DEWEY, JOSEPH C. "Consistency of Pupil Response on Tests of Reading Comprehension." *Elementary School Journal* 34: 533-36; March, 1934.
293. DICKEY, JOHN W. "On Estimating the Reliability Coefficient." *Journal of Applied Psychology* 18: 103-15; February, 1934.
294. DUNLAP, JACK W. "Comparable Tests and Reliability." *Journal of Educational Psychology* 24: 442-53; September, 1933.
295. ELBERTON, MARION. "An Experiment in Map Scoring and Mental Imagery Tests." *Journal of Applied Psychology* 17: 376-406; August, 1933.
296. FAY, PAUL J., and MIDDLETON, WARREN C. "An Economical Method of Administering and Scoring New-Type Examinations." *Journal of Applied Psychology* 18: 77-84; February, 1934.

297. FRUTCHHEY, FRED P. "Measuring the Ability to Apply Chemical Principles." *Educational Research Bulletin (Ohio State University)* 12: 255-60; December 13, 1933.
298. GERBERICH, J. R. "A Technique for Measuring the Ability to Evaluate Objective Test Items." *Journal of Educational Research* 27: 46-50; September, 1933.
299. HANDY, UVAN, and LENTZ, THEODORE F. "Item Value and Test Reliability." *Journal of Educational Psychology* 25: 703-8; December, 1934.
300. HENRY, LORNE J. "A Comparison of the Difficulty and Validity of Achievement Test Items." *Journal of Educational Psychology* 25: 537-41; October, 1934.
301. HEVNER, KATE. "A Method of Correcting for Guessing in True-False Tests and Empirical Evidence in Support of It." *Journal of Social Psychology* 3: 359-62; August, 1932.
302. HOLZINGER, KARL J. "The Reliability of a Single Test Item." *Journal of Educational Psychology* 23: 411-17; September, 1932.
303. HOLZINGER, KARL J., and SWINEFORD, FRANCES, compilers. "Selected References on Statistics and the Theory of Test Construction." *School Review* 41: 462-66; June, 1933. 42: 459-65; June, 1934. 43: 462-67; June, 1935.
304. HORN, HELEN R. "The Variable Answer Test." *English Journal (H. S. Ed.)* 23: 223-25; March, 1934.
305. HORST, PAUL. "The Difficulty of a Multiple-Choice Test Item." *Journal of Educational Psychology* 24: 229-32; March, 1933.
306. HORST, PAUL. "The Economical Collection of Data for Test Validation." *Journal of Experimental Education* 2: 250-53; March, 1934.
307. HORST, PAUL. "Item Analysis by the Method of Successive Residuals." *Journal of Experimental Education* 2: 254-63; March, 1934.
308. HURD, A. W. "Comparisons of Short Answer and Multiple Choice Tests Covering Identical Subject Content." *Journal of Educational Research* 26: 28-30; September, 1932.
309. KELLEY, TRUMAN L. "The Scoring of Alternative Responses with Reference to Some Criterion." *Journal of Educational Psychology* 25: 504-10; October, 1934.
310. KELLEY, VICTOR H. "An Experimental Study of Certain Techniques for Testing Word Meanings." *Journal of Educational Research* 27: 277-82; December, 1933.
311. KEYS, NOEL. "The Influence of True-False Items on Specific Learning." *Journal of Educational Psychology* 25: 511-20; October, 1934.
312. KRUEGER, WM. C. F. "Distributions of Scores Based on Correct Guessing for True-False Tests of Various Lengths." *Journal of Educational Psychology* 24: 185-88; March, 1933.
313. LAMSON, EDNA E. "What Happens When the Second Judgment is Recorded in a True-False Test?" *Journal of Educational Psychology* 26: 223-27; March, 1935.
314. LEE, J. MURRAY, and SYMONDS, PERCIVAL M. "New Type or Objective Tests: A Summary of Recent Investigations. (October, 1931-October, 1933)." *Journal of Educational Psychology* 25: 161-84; March, 1934.
315. LINDQUIST, E. F., and ANDERSON, H. R. "Achievement Tests in the Social Studies." *Educational Record* 14: 198-256; April, 1933.
316. LINDQUIST, E. F., and COOK, WALTER W. "Experimental Procedures in Test Evaluation." *Journal of Experimental Education* 1: 163-85; March, 1933.
317. LONG, JOHN A. "Improved Overlapping Methods for Determining Validities of Test Items." *Journal of Experimental Education* 2: 264-68; March, 1934.
318. McCLUSKY, HOWARD YAK. "An Experimental Comparison of Two Methods of Correcting the Outcome of an Examination." *School and Society* 40: 566-68; October 27, 1934.
319. McCLUSKY, HOWARD YAK. "The Negative Suggestion Effect of the False Statement in the True-False Test." *Journal of Experimental Education* 2: 269-73; March, 1934.
320. MACILL, WALTER H. "The Influence of the Form of Item on the Validity of Achievement Tests." *Journal of Educational Psychology* 25: 21-28; January, 1934.
321. MELBO, IRVING R. "How Much Do Students Guess in Taking True-False Examinations?" *Educational Method* 12: 485-87; May, 1933.
322. MEYER, GEORGE. "An Experimental Study of the Old and New Types of Examination: I. The Effect of the Examination Set on Memory." *Journal of Educational Psychology* 25: 641-61; December, 1934.

323. MEYER, GEORGE. "An Experimental Study of the Old and New Types of Examination: II. Methods of Study." *Journal of Educational Psychology* 26: 30-40; January, 1935.
324. MINNESOTA UNIVERSITY. COMMITTEE ON EDUCATIONAL RESEARCH. *Studies in College Examinations*. Minneapolis: the University, 1934. 204 p.
325. MONROE, WALTER S. "Hazards in the Measurement of Achievement." *School and Society* 41: 48-52; January 12, 1935.
326. PRESSEY, S. L. "A Third and Fourth Contribution toward the Coming 'Industrial Revolution' in Education." *School and Society* 36: 668-72; November 19, 1932.
327. PRICE, ROY A. "Tests in the Social Studies." *Social Studies* 26: 23-29; January, 1935.
328. RAUTH, J. EDWARD. "Scoring Objective Tests." *Catholic Educational Review* 33: 140-47; March, 1935.
329. RICHARDSON, M. W., and STALNAKER, JOHN M. "A Note on the Use of Biserial r in Test Research." *Journal of General Psychology* 8: 463-65; April, 1933.
330. ROSS, ROBERT T., and PRIE, MARJORIE. "The Persistence of Errors in Successive True-False Tests." *Journal of Educational Psychology* 25: 422-26; September, 1934.
331. SCHEIDEMANN, NORMA V. "Multiplying the Possibilities of the Multiple Choice Form of Objective Question." *Journal of Applied Psychology* 17: 337-40; June, 1933.
332. SIMS, VERNER MARTIN. "An Evaluation of Five-, Ten-, and Fifteen-Item Rearrangement Tests." *Journal of Educational Psychology* 25: 251-57; April, 1934.
333. SIMS, VERNER MARTIN, and KNOX, L. B. "The Reliability and Validity of Multiple-Response Tests When Presented Orally." *Journal of Educational Psychology* 23: 656-62; December, 1932.
334. SMELTZER, C. H. "Objective Measurement of Applied Information." *Journal of Applied Psychology* 17: 765-71; December, 1933.
335. SMITH, MAX. *The Relationship between Item Validity and Test Validity*. Contributions to Education, No. 621. New York: Teachers College, Columbia University, 1934. 40 p. (Summarized in *Teachers College Record* 36: 146-47; November, 1934.)
336. SPROULE, CHESTER E. "Suggestion Effects of the True-False Test." *Journal of Educational Psychology* 25: 281-85; April, 1934.
337. STALNAKER, JOHN M., and RUTH C. "Chance versus Selected Distractors in a Vocabulary Test." *Journal of Educational Psychology* 26: 161-68; March, 1935.
338. STEPHENSON, W. "Factorizing the Reliability Coefficient." *British Journal of Psychology (General Section)* 25: 211-16; October, 1934.
339. TERRY, PAUL W. "How Students Review for Objective and Essay Tests." *Elementary School Journal* 33: 592-603; April, 1933.
340. TERRY, PAUL W. "How Students Study for Three Types of Objective Tests." *Journal of Educational Research* 27: 333-43; January, 1934.
341. THURSTONE, L. L. *The Reliability and Validity of Tests*. Ann Arbor, Mich.: Edwards Brothers, 1931. 113 p.
342. TINKER, MILES A. "Influence of the Speed Attitude on Test Performance." *Journal of General Psychology* 10: 465-69; April, 1934.
343. TURNEY, AUSTIN H. "The Concept of Validity in Mental and Achievement Testing." *Journal of Educational Psychology* 25: 81-95; February, 1934.
344. TYLER, RALPH W. "Assumptions Involved in Achievement-Test Construction." *Educational Research Bulletin (Ohio State University)* 12: 29-36; February 8, 1933.
345. TYLER, RALPH W. "The Construction of Examinations in Botany and Zoölogy." *Service Studies in Higher Education*, Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 43-51.
346. TYLER, RALPH W. "Formulating Objectives for Tests." *Educational Research Bulletin (Ohio State University)* 12: 197-206; October 11, 1933.
347. TYLER, RALPH W. "Techniques for Evaluating Behavior." *Educational Research Bulletin (Ohio State University)* 13: 1-11; January 17, 1934.
348. VOTAW, DAVID F. "Graphical Determination of Probable Error in Validation of Test Items." *Journal of Educational Psychology* 24: 682-86; December, 1933.
349. VOTAW, DAVID F. "Notes on Validation of Test Items by Comparison of Widely Spaced Groups." *Journal of Educational Psychology* 25: 185-91; March, 1934.

350. WATSON, GOODWIN. "Note on Validity in the Measurement of Change." *Journal of Educational Research* 27: 187-92; November, 1933.
351. WEIDEMANN, CHARLES C., and NEWENS, LYNDALE FISHER. "The Effect of Directions Preceding True-False and Indeterminate Statement Examinations upon Distributions of Test Scores." *Journal of Educational Psychology* 24: 97-106; February, 1933.
352. WELLER, LOUISE, and BROOM, M. E. "A Study of the Validity of Six Types of Spelling Tests." *School and Society* 40: 103-4; July 21, 1934.
353. WILLOUGHBY, RAYMOND ROYCE. "The Concept of Reliability." *Psychological Review* 42: 153-65; March, 1935.
354. WOODY, CLIFFORD, and OTHERS. "A Symposium on the Effects of Measurement on Instruction." *Journal of Educational Research* 28: 481-527; March, 1935.
355. WORCESTER, D. A. "On the Validity of Testing." *School Review* 42: 527-31; September, 1934.
356. ZUBIN, JOSEPH. "The Chance Element in Matching Tests." *Journal of Educational Psychology* 24: 674-81; December, 1933.
357. ZUBIN, JOSEPH. "The Method of Internal Consistency for Selecting Test Items." *Journal of Educational Psychology* 25: 345-56; May, 1934.

Chapter IV. Recent Developments in the Written Essay Examination

358. CALDWELL, FLOYD F., and MOWRY, MARY DAVIS. "The Essay versus the Objective Examination as Measures of the Achievement of Bilingual Children." *Journal of Educational Psychology* 24: 696-702; December, 1933.
359. COCHRAN, ROY E., and WEIDEMANN, CHARLES C. *The 'Discuss' Essay versus the Word-Answer Fact Test*. Unpublished manuscript.
360. COCHRAN, ROY E., and WEIDEMANN, CHARLES C. "'Explain' Essay versus Word-Answer Fact Test." *Phi Delta Kappan* 18: 59-61, 75; December, 1934.
361. COCHRAN, ROY E., and WEIDEMANN, CHARLES C. *Improvement of the Consistency of Scoring the 'Explain' and 'Discuss' Essay Tests*. Unpublished manuscript.
362. COREY, S. M. "The Correlation between New Type and Essay Examination Scores, and the Relationship between Them and Intelligence as Measured by Army Alpha." *School and Society* 32: 849-50; December 20, 1930.
363. CRAWFORD, C. C. "How To Study for Objective Test." *Education* 53: 413-16; March, 1933.
364. CRAWFORD, C. C., and RAYNALDO, D. A. "Some Experimental Comparisons of True-False Tests and Traditional Examinations." *School Review* 33: 698-706; November, 1925.
365. DOUGLASS, HARL R., and TALLMADGE, MARGARET. "How University Students Prepare for New Types of Examinations." *School and Society* 39: 318-20; March 10, 1934.
366. EELLS, W. C. "Reliability of Repeated Grading of Essay Type Examinations." *Journal of Educational Psychology* 21: 48-52; January, 1930.
367. HASTINGS, HARRY WORTHINGTON. "Student-Made Examination." *English Journal* (Coll. Ed.) 21: 151-53; February, 1932.
368. HUDELSON, EARL. "The Effect of Objective Standards upon Composition Teachers' Judgments." *Journal of Educational Research* 12: 329-40; December, 1925.
369. KINNEY, L. B., and EURICH, A. C. "A Summary of Investigations Comparing Different Types of Tests." *School and Society* 36: 540-44; October 22, 1932.
370. KLISE, NIRA M. "Student Opinions of Type of Examination." *School and Society* 24: 23-24; July 3, 1926.
371. LEIGHTON, R. W. "Improvement of the Essay Type Examination." *Research in Higher Education*. U. S. Dept. of the Interior, Office of Education, Bulletin, 1931, No. 12. Washington, D. C.: Government Printing Office, 1932. p. 15-20.
372. MCKEE, JAMES HUGH. "Subjective and (or versus) Objective." *English Journal* (Coll. Ed.) 23: 127-33; February, 1934.
373. MEYER, GEORGE. "The Essay Type of Examination." *American School Board Journal* 89: 17-18; December, 1934.
374. MEYER, GEORGE. "Experimental Study of the Old and New Types of Examination. I." *Journal of Educational Psychology* 25: 641-61; December, 1934.

375. MONROE, WALTER S., and SOUDERS, LLOYD B. *The Present Status of Written Examinations and Suggestions for Their Improvement*. University of Illinois Bureau of Educational Research Bulletin No. 17. Urbana, Ill.: the University, 1923. 77 p.
376. ODELL, C. W. *The Use of Scales for Rating Pupils' Answers to Thought Questions*. University of Illinois Bureau of Educational Research Bulletin No. 46. Urbana, Ill.: the University, 1929. 34 p.
377. OSBURN, WORTH J. "Testing Thinking." *Journal of Educational Research* 27: 401-11; February, 1934.
378. PATERSON, DONALD G. "Do New and Old Type Examinations Measure Different Mental Functions?" *School and Society* 24: 246-48; August 21, 1926.
379. PETERS, C. C., and MARTZ, H. B. "A Study of the Validity of Various Types of Examinations." *School and Society* 33: 336-38; March 7, 1931.
380. RUCH, G. M., and STODDARD, GEORGE D. *Tests and Measurements in High School Instruction*. Yonkers-on-Hudson, N. Y.: World Book Co., 1927. 381 p.
381. SIMS, VERNER MARTIN. "Essay Examination Questions Classified on the Basis of Objectivity." *School and Society* 35: 100-2; January 16, 1932.
382. SIMS, VERNER MARTIN. "Improving the Measuring Qualities of an Essay Examination." *Journal of Educational Research* 27: 20-31; September, 1933.
383. SIMS, VERNER MARTIN. "The Objectivity, Reliability and Validity of an Essay Examination Graded by Rating." *Journal of Educational Research* 24: 216-23; October, 1931.
384. SIMS, VERNER MARTIN. "Reducing the Variability of Essay Examination Marks through Eliminating Variations in Standards of Grading." *Journal of Educational Research* 26: 637-47; May, 1933.
385. TERRY, PAUL W. "How Students Review for Objective and Essay Tests." *Elementary School Journal* 33: 592-603; April, 1933.
386. THARP, JAMES BURTON. "The New Examination versus the Old in Foreign Languages." *School and Society* 26: 691-94; November 26, 1927.
387. WEAVER, ROBERT B., and TRAXLER, ARTHUR E. "Essay Examinations and Objective Tests in United States History in the Junior High School." *School Review* 39: 689-95; November, 1931.
388. WEIDEMANN, CHARLES C., and NEWENS, LYNDALE FISHER. "Does the Compare-and-Contrast Essay Test Measure the Same Mental Functions as the True-False Test?" *Journal of General Psychology* 9: 430-49; October, 1933.
389. WOOD, BEN D. "The Measurement of Law School Work." *Columbia Law Review* 24: 221-65; March, 1924. 25: 316-31; March, 1925.

Chapter V. Achievement Tests in Colleges and Universities

390. AMERICAN COUNCIL ON EDUCATION, COMMITTEE ON PERSONNEL METHODS. *Measurement and Guidance of College Students*. Baltimore: Williams and Wilkins Co., 1933. 199 p.
391. BARR, A. S. "The Effects of Measurement on Instruction." *Journal of Educational Research* 28: 481-83; March, 1935.
392. BARR, A. S. "The Measurement of Teaching Ability." *Journal of Educational Research* 28: 561-69; April, 1935.
393. BETTS, GEORGE H. "General Information Possessed by Graduate Students in Education." *School and Society* 37: 821-24; June 24, 1933.
394. BOUCHER, C. S. *The Chicago College Plan*. Chicago: University of Chicago Press, 1935. 344 p.
395. BOUCHER, C. S. "Examinations at the University of Chicago." *Bulletin of the Association of American Colleges* 21: 103-7; March, 1935.
396. BOWDEN, A. O. "Change—The Test of Teaching." *School and Society* 40: 133-36; July 28, 1934.
397. BRIGHAM, CARL C. "Admission Units and Freshman Placement." *Educational Record* 15: 56-67; January, 1934.
398. BROWNELL, WILLIAM A. "The Use of Objective Measures in Evaluating Instruction." *Educational Method* 13: 401-8; May-June, 1934.

399. CARROLL, HERBERT A. "A Method of Measuring Prose Appreciation." *English Journal* 22: 184-89; March, 1933.
400. CHARTERS, W. W. "Education and Research at a Mechanics Institute: A Character Development Study." *Personnel Journal* 12: 119-23; August, 1933.
401. CHEYDLEUR, F. D. "An Experiment in Adult Learning of French at the Madison, Wisconsin, Vocational School." *Journal of Educational Research* 26: 259-75; December, 1932.
402. CRAWFORD, ALBERT BEECHER. "Some Criticisms of Current Practice in Educational Measurements." *Harvard Teachers Record* 3: 67-81; April, 1933.
403. DOUGLASS, HARL R. "The Effect of Measurement upon Instruction." *Journal of Educational Research* 28: 508-11; March, 1935.
404. DOUGLASS, HARL R. "Some Dangers of the Testing Movement." *Journal of the National Education Association* 23: 17-18; January, 1934.
405. ELLIS, ELMER. "The Permanence of Learning in World History." *Social Studies* 25: 133-36; March, 1934.
406. EURICH, ALVIN C. "Measuring the Achievement of Objectives in Freshmen English." *Studies in College Examinations*. Minneapolis: University of Minnesota Press, 1934. p. 51-66.
407. FLETCHER, HARRIS F. "Proficiency Examinations for Credit at the University of Illinois." *School and Society* 36: 792-93; December 17, 1932.
408. GAUSEWITZ, WALTER. "Adapting Elementary German to the Needs of the Gifted Student." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 135-40.
409. HAGGERTY, M. E. "Product of Higher Educational Institutions." *North Central Association Quarterly* 8: 248-61; September, 1933.
410. HANFORD, A. C. "The General Examinations in Harvard College." *Bulletin of the Association of American Colleges* 21: 107-14; March, 1935.
411. HARTMANN, GEORGE W., and BARRICK, FLOYD M. "Fluctuations in General Cultural Information among Undergraduates." *Journal of Educational Research* 28: 255-64; December, 1934.
412. HARTMANN, GEORGE W. *Measuring Teaching Efficiency among College Instructors*. Archives of Psychology, No. 154. New York: Columbia University, 1933. 45 p.
413. HENDRICKS, B. C.; TYLER, R. W.; and FRUTCHEY, F. P. "Testing Ability to Apply Chemical Principles." *Journal of Chemical Education* 11: 611-13; November, 1934.
414. HENMON, V. A. C. "Recent Developments in the Study of Modern Foreign Language Problems." *Modern Language Journal* 19: 187-201; December, 1934.
415. HORNEY, AMOS G. "Testing the Achievement of Students in Chemistry." *Journal of Chemical Education* 11: 360-66; June, 1934.
416. HORST, PAUL. "Increasing the Efficiency of Selection Tests." *Personnel Journal* 12: 254-59; February, 1934.
417. JOHNSON, PALMER O. "A Measurement Program in Junior College Science." *Science Education* 17: 176-82; October, 1933.
418. JOHNSTON, J. B., chairman. "The 1933 College Sophomore Testing Program." *Educational Record* 14: 522-71; October, 1933.
419. JOHNSTON, J. B., chairman. "The 1934 College Sophomore Testing Program." *Educational Record* 15: 471-516; October, 1934.
420. JONES, EDWARD SAFFORD. *Comprehensive Examinations in American Colleges*. An investigation for the Association of American Colleges. New York: Macmillan Co., 1933. 436 p.
421. KEELER, L. W. "Measurement and Instruction." *Journal of Educational Research* 28: 493-95; March, 1935.
422. KELLEY, TRUMAN L., and KREY, A. C. *Tests and Measurements in the Social Sciences*. Report of the Commission on the Social Studies of the American Historical Association, Part IV. New York: Charles Scribner's Sons, 1934. 635 p.
423. KREY, A. C. "The Effect of Measurement on Instruction." *Journal of Educational Research* 28: 498-501; March, 1935.
424. KULP, DANIEL H. "Weekly Tests for Graduate Students?" *School and Society* 38: 157-59; July 29, 1933.

425. LEARNED, WILLIAM S. "Study of the Relations of Secondary and Higher Education in Pennsylvania: Knowledge as a Factor in Education—The Tests and Their Implications." *Twenty-eighth Annual Report of the President and of the Treasurer*. New York: Carnegie Foundation for the Advancement of Teaching, 1933. p. 39-63.
426. LEHMAN, HARVEY C., and WITTY, PAUL A. "Objectives and Aims in the Introductory Course in Psychology." *Journal of Applied Psychology* 18: 681-95; October, 1934.
427. LINDQUIST, E. F., and ANDERSON, H. R. "Achievement Tests in the Social Studies." *Educational Record* 14: 198-256; April, 1933.
428. LINDQUIST, E. F., and COOK, WALTER W. "Experimental Procedures in Test Evaluation." *Journal of Experimental Education* 1: 163-85; March, 1933.
429. LOWELL, A. LAWRENCE. "Examination in the Educational Process." *Harvard Teachers Record* 3: 184-88; October, 1933.
430. MEYER, BLANCHE B. M. "Remedial Instruction for Students Having Difficulty in Zoölogy." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 93-108.
431. MILLER, DAVID F. "Special Treatment for Superior Students in General Zoölogy." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 72-78.
432. MONROE, ROBERT E. "Adapting Instruction to the Ability of the Student in the Romance Languages." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 123-34.
433. MONROE, WALTER S. "Effect of Measurement on Instruction." *Journal of Educational Research* 28: 496-97; March, 1935.
434. MONROE, WALTER S. "Hazards in the Measurement of Achievement." *School and Society* 41: 48-52; January 12, 1935.
435. NOLL, VICTOR H. "Measuring Scientific Thinking." *Teachers College Record* 35: 685-93; May, 1934.
436. PALMER, FREDERIC. "The College Physics Testing Program and Its Significance for Guidance in Secondary Schools." *Educational Record* 16: 82-96; January, 1935.
437. PRICE, JOHN W., and MILLER, JOHN A. "An Experiment in Sectioning Students in the Second Course in Zoölogy." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 79-92.
438. REEDER, CHARLES W. "Forecasting Academic Success in the College of Commerce and Administration." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 206-20.
439. RICHARDSON, M. W., and STALNAKER, JOHN M. "Comments on Achievement Examinations." *Journal of Educational Research* 28: 425-32; February, 1935.
440. ROBBINS, WILLIAM J. "The Coordination and Standardization of Means Other Than Examinations for Expediting and Measuring a Student's Achievement." *Journal of Proceedings of the Association of American Universities*, 1933. Chicago: University of Chicago Press, 1933. p. 64-66.
441. SINCLAIR, J. H., and TOLMAN, R. S. "Attempt to Study the Effect of Scientific Training upon Prejudice and Illogicality of Thought." *Journal of Educational Psychology* 24: 362-70; May, 1933.
442. SMITH, MAX. *The Relationship Between Item Validity and Test Validity*. Contributions to Education, No. 621. New York: Teachers College, Columbia University, 1934. 40 p.
443. STALNAKER, JOHN M., and STALNAKER, RUTH C. "Open-Book Examinations." *Journal of Higher Education* 5: 117-20; March, 1934.
444. STALNAKER, JOHN M., and STALNAKER, RUTH C. "Open-Book Examinations: Results." *Journal of Higher Education* 6: 214-16; April, 1935.
445. STALNAKER, JOHN M., and STALNAKER, RUTH C. "Reliable Reading of Essay Tests." *School Review* 42: 599-605; October, 1934.
446. STALNAKER, JOHN M., and RICHARDSON, M. W. "Scholarship Examinations." *Journal of Higher Education* 5: 305-13; June, 1934.

447. TERRY, PAUL W. "The Prognostic Value of Different Types of Tests in Courses in Educational Psychology." *Journal of Applied Psychology* 18: 231-40; April, 1934.
448. THARP, JAMES B. "A Modern Language Test." *Journal of Higher Education* 6: 103-4; February, 1935.
449. THEISEN, W. W. "The Effects of Measurement on Instruction." *Journal of Educational Research* 28: 484-87; March, 1935.
450. TRIMBLE, OTIS C. "The Oral Examination: Its Validity and Reliability." *School and Society* 39: 550-52; April 28, 1934.
451. TURNBULL, EVE E., and GRIFFITH, MARION E. "Revision of the Elementary Courses in Textiles." *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. p. 169-86.
452. TYLER, HENRY T. "Remedial Reading in the Junior College." *Junior College Journal* 4: 28-31; October, 1933.
453. TYLER, RALPH W. *Constructing Achievement Tests*. Columbus: Ohio State University, 1934. 102 p. (Reprints from *Educational Research Bulletin*.)
454. TYLER, RALPH W., and OTHERS. *Service Studies in Higher Education*. Bureau of Educational Research Monographs, No. 15. Columbus: Ohio State University, 1932. 283 p.
455. UHL, WILLIS L. "Some Neglected Aspects of Educational Measurement." *Journal of Educational Research* 27: 241-46; December, 1933.
456. WAGNER, MAZIE E., and STRABEL, EUNICE. "Predicting Success and Failure in College Ancient and Modern Foreign Languages." *Modern Language Journal* 19: 285-93; January, 1935.
457. WELBORN, ERNEST L. "A Study of Logical Learning in College Classes." *Twentieth Annual Conference on Educational Measurements*. School of Education Bulletin, Vol. 10, No. 1. Bloomington, Indiana: Indiana University, 1933. p. 12-20.
458. WILSON, GUY M. "Tests: Aims, Procedure and Types." *Journal of Educational Research* 28: 488-92; March, 1935.
459. WOOD, BEN D. "Major Strategy versus Minor Tactics in Educational Testing." *Baltimore Bulletin of Education* 13: 3-16; September, 1934.
460. WOOD, ELEANOR PERRY "Examining the Uses of Examinations." *Harvard Teachers Record* 3: 59-66; April, 1933.
461. WOODY, CLIFFORD. "Summary and Reactions." *Journal of Educational Research* 28: 520-27; March, 1935.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

Membership¹

HONORARY AND LIFE

- Ayres, Leonard P., Vice President, Cleveland Trust Co., Cleveland, Ohio.
Buckingham, B. R., Editorial Department, Ginn and Co., 15 Ashburton Place, Boston, Massachusetts.
Cattell, J. McKeen, Editor of *Science and School and Society*, Garrison, New York.
Coffman, L. D., President, University of Minnesota, Minneapolis, Minnesota.
Hanus, Paul H., Professor of Education Emeritus, Harvard University, Cambridge, Massachusetts. (3 Channing Circle, Cambridge, Massachusetts.)
Judd, Charles H., Head of Department of Education, University of Chicago, Chicago, Illinois.
Russell, James E., Dean Emeritus, Teachers College, Columbia University, New York, New York. (R. F. D. 4, Trenton, New Jersey.)
Russell, William F., Dean, Teachers College, Columbia University, New York, New York.
Terman, Lewis M., Professor of Psychology, Stanford University, California.
Thorndike, E. L., Professor of Education, Columbia University, New York, New York.
Wissler, Clark, Professor of Anthropology, Institute of Human Relations, Yale University, New Haven, Connecticut.
Zook, George F., President, American Council on Education, 744 Jackson Place, N. W., Washington, D. C.

ACTIVE

- Ade, Lester Kelly, State Superintendent of Public Instruction, Harrisburg, Pennsylvania.
Alexander, Carter, Library Professor, Teachers College, Columbia University, New York, New York.
Allen, Ira M., Superintendent of Schools, Highland Park, Michigan.
Alschuler, Mrs. Rose H., Staff Director, Winnetka Public School Nursery Unit, Skokie School, Winnetka, Illinois.
Alves, H. F., Senior Specialist in State School Administration, United States Office of Education, Washington, D. C.
Anderson, Earl W., Professor of Education, Bureau of Educational Research, Ohio State University, Columbus, Ohio.
Andrus, Ruth, Chief of Child Development and Parent Education Bureau, State Education Department, Albany, New York.
Arnold, William E., Assistant Professor of Education, University of Pennsylvania, Philadelphia, Pennsylvania.
Ashbaugh, E. J., Dean, School of Education, Miami University, Oxford, Ohio.
Averill, William A., Instructor in Education, Lesley Normal School, Cambridge, Massachusetts.
Ayer, Fred C., Professor of Educational Administration, University of Texas, Austin, Texas.
Baer, Joseph A., Administrative Associate, Ethical Culture Schools, New York, New York.
Baker, Harry J., Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.
Bamberger, Florence E., Professor of Education, Johns Hopkins University, Baltimore, Maryland.
Barr, A. S., Professor of Education, University of Wisconsin, Madison, Wisconsin.

¹ Corrected up to December 1, 1935. Errors should be reported to the Secretary-Treasurer immediately.

- Barthelmess, Harriet M.**, Division of Educational Research, Board of Education, Philadelphia, Pennsylvania.
- Barton, W. A., Jr.**, Head, Department of Education, Psychology, and Philosophy, Coker College, Hartsville, South Carolina.
- Beeby, C. E.**, Chief Executive Officer, New Zealand Council for Educational Research, Southern Cross Building, Wellington, C. 1, New Zealand.
- Benjamin, Harold**, Professor of Education, University of Minnesota, Minneapolis, Minnesota.
- Benz, H. E.**, Associate Professor of Mathematics, College of Education, Ohio University, Athens, Ohio.
- Bergman, W. G.**, Department of Research, Detroit Public Schools, Detroit, Michigan.
- Betts, Gilbert L.**, Director of the Curriculum Department, West Allis Public Schools, West Allis, Wisconsin.
- Billett, Roy O.**, Professor of Education, Boston University, Boston, Massachusetts.
- Bixler, Harold H.**, Director of Research and Guidance, Board of Education, City Hall, Atlanta, Georgia.
- Boardman, Charles W.**, Professor of Education, University of Minnesota, Minneapolis, Minnesota.
- Booker, Ivan A.**, Assistant Director, Research Division, National Education Association, Washington, D. C.
- Bowyer, Vernon**, Administrative Assistant, Emergency Education Program, Board of Education, Chicago, Illinois.
- Boyer, Philip A.**, Director, Division of Educational Research, Board of Education, Administration Building, Philadelphia, Pennsylvania.
- Brainerd, Mrs. Margaret**, 1101 Walnut Street, Martins Ferry, Ohio.
- Branson, Ernest P.**, Counselor, Polytechnic High School, Long Beach, California.
- Breed, Frederick S.**, Associate Professor of Education, University of Chicago, Chicago, Illinois.
- Bright, Ira J.**, Superintendent of Schools, Leavenworth, Kansas.
- Bristow, William H.**, Director, Bureau of School Curriculum, State Department of Public Instruction, Harrisburg, Pennsylvania.
- Broening, Angela M.**, Head of the English Department, Forest Park High School, and Chairman, Curriculum Research in Secondary English, Public Schools, Baltimore, Maryland.
- Brooks, Fowler D.**, Head, Departments of Education and Psychology, De Pauw University, Greencastle, Indiana.
- Brown, Clara M.**, Associate Professor of Home Economics, University Farm, St. Paul, Minnesota.
- Brown, Edwin J.**, Director, Graduate Division, Kansas State Teachers College, Emporia, Kansas.
- Brownell, S. M.**, Superintendent of Schools, Grosse Point, Michigan.
- Brownell, W. A.**, Professor of Educational Psychology, Duke University, Durham, North Carolina.
- Brueckner, Leo J.**, Professor of Elementary Education, University of Minnesota, Minneapolis, Minnesota.
- Brumbaugh, A. J.**, Dean of Students in the College and Assistant Professor of Education, University of Chicago, Chicago, Illinois.
- Bruner, H. B.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Brunner, Edmund de S.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Buckner, C. A.**, Professor of Education, University of Pittsburgh, Pittsburgh, Pennsylvania.
- Buros, Francis C.**, Executive Assistant to the Superintendent of Schools, White Plains, New York.
- Buros, Oscar K.**, Assistant Professor of Education, Rutgers University, New Brunswick, New Jersey.
- Burr, Samuel Engle**, Superintendent of Schools, New Castle, Delaware.
- Bursch, James F.**, Director of Research and Student Personnel, Sacramento City Schools, Sacramento, California.
- Buswell, G. T.**, Professor of Educational Psychology, University of Chicago, Chicago, Illinois.

Butsch, R. L. C., Associate Professor of Education, Graduate School, Marquette University, Milwaukee, Wisconsin.

Butterworth, Julian E., Director, Graduate School of Education, Cornell University, Ithaca, New York.

Caldwell, Otis W., Director, Institute of School Experimentation and Professor of Education, Teachers College, Columbia University, New York, New York.

Cammack, James W., Jr., Director of Research, State Department of Education, Frankfort, Kentucky.

Campbell, Doak S., Professor of Education, Division of Surveys and Field Studies, George Peabody College for Teachers, Nashville, Tennessee.

Carr, William G., Director, Research Division, National Education Association, Washington, D. C.

Carroll, Herbert A., Assistant Professor of Educational Psychology, University of Minnesota, Minneapolis, Minnesota.

Caswell, Hollis L., Professor of Education, Division of Surveys and Field Studies, George Peabody College for Teachers, Nashville, Tennessee.

Cattell, Psyche, Research Fellow, School of Public Health, Palfrey House, Harvard University, Cambridge, Massachusetts.

Cavins, L. V., Director of Research, State Department of Education, Charleston, West Virginia.

Chambers, M. M., Staff Member, American Youth Commission, 744 Jackson Place, N. W., Washington, D. C.

Chapman, Harold B., Assistant Director, Bureau of Educational Research, Public Schools, Baltimore, Maryland.

Charters, W. W., Director, Bureau of Educational Research, Ohio State University, Columbus, Ohio.

Chase, Vernon Emory, Director, Bureau of Research and Adjustment, Public Schools, Dearborn, Michigan.

Clapp, Frank L., Professor of Education, University of Wisconsin, Madison, Wisconsin.

Clark, Zenas R., Director of Research, Wilmington Public Schools, Wilmington, Delaware.

Clem, Orlie M., Superintendent of Schools, Owego, Tioga County, New York.

Cobb, Margaret V., Tilton, New Hampshire.

Cocking, Walter D., State Commissioner of Education, Nashville, Tennessee.

Coffey, Wilford L., Research Student, Teachers College, Columbia University, New York, New York.

Connor, William L., Chief, Bureau of Educational Research, Board of Education, Cleveland, Ohio.

Conrad, Herbert S., Assistant Professor of Education, Department of Education, University of California, Berkeley, California.

Cooke, Dennis H., Professor of School Administration, George Peabody College for Teachers, Nashville, Tennessee.

Coon, Beulah I., Agent for Studies and Research in Home Economics Education, United States Office of Education, Washington, D. C.

Cornell, Ethel L., Research Associate, Educational Research Division, State Education Department, Albany, New York.

Couns, George S., Professor of Education, Teachers College, Columbia University, New York, New York.

Courtis, S. A., Professor of Education, University of Michigan, Ann Arbor, Michigan.

Coxe, W. W., Director, Educational Research Division, State Department of Education, Albany, New York.

Coy, Genevieve L., Psychologist, Dalton Schools, New York, New York.

Craig, Gerald S., Associate Professor, Natural Sciences, Teachers College, Columbia University, New York, New York.

Crawford, C. C., Professor of Education, University of Southern California, Los Angeles, California.

Cutright, Prudence, Assistant Superintendent, Minneapolis Public Schools, Minneapolis, Minnesota.

Cutts, Norma E., Supervisor, Department of Exceptional Children, Board of Education, New Haven, Connecticut.

Dale, Edgar, Associate Professor, College of Education, Ohio State University, Columbus, Ohio.

- Davis, Mary Dabney**, Specialist, United States Office of Education, Washington, D. C.
- Dawson, Howard A.**, Assistant Director, Research Division, National Education Association, Washington, D. C.
- Dearborn, Ned H.**, Dean, Division of General Education, New York University, New York, New York.
- Deffenbaugh, Walter S.**, Chief, American School Systems Division, United States Office of Education, Washington, D. C.
- DeLong, Leo R.**, Supervising Principal, Raritan Township, Stelton, New Jersey.
- DeVoss, J. C.**, Dean, Upper Division, San Jose State College, San Jose, California.
- Dolch, E. W.**, Assistant Professor of Education, University of Illinois, Urbana, Illinois.
- Douglass, Harl R.**, Professor of Education, University of Minnesota, Minneapolis, Minnesota.
- Downing, Elliot R.**, Associate Professor Emeritus, the Teaching of Science, University of Chicago, Chicago, Illinois. (P. O. Box 147, Williams Bay, Wisconsin.)
- Durost, Walter N.**, Test Editor, World Book Company, Yonkers, New York.
- Durrell, Donald D.**, Professor of Education, Boston University, Boston, Massachusetts.
- Eads, Laura Krieger**, Research Associate, Erpi Pictures Consultants, Inc., New York, New York.
- Easley, Howard**, Assistant Professor of Educational Psychology, Duke University, Durham, North Carolina.
- Edmiston, Robert Wentz**, Director of Extension, Miami University, Oxford, Ohio.
- Edmonson, James B.**, Dean, School of Education, University of Michigan, Ann Arbor, Michigan.
- Edwards, Newton**, Professor of Education, University of Chicago, Chicago, Illinois.
- Eells, Walter C.**, Coordinator, Cooperative Study of Secondary School Standards, 744 Jackson Place, N. W., Washington, D. C.
- Elliott, Charles H.**, Commissioner of Education for New Jersey, Trenton, New Jersey.
- Elliott, Eugene B.**, State Superintendent of Public Instruction, Lansing, Michigan.
- Elsbree, Willard S.**, Associate Professor of Education, Teachers College, Columbia University, New York, New York.
- Engelhardt, N. L.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Eurich, Alvin C.**, Assistant Director, Educational Research, University of Minnesota, Minneapolis, Minnesota.
- Evenden, Edward S.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Feder, Daniel D.**, Associate, Psychology and Personnel, University of Iowa, Iowa City, Iowa.
- Ferriss, Emery N.**, Professor of Education, Rural Educational Department, Cornell University, Ithaca, New York.
- Flemming, Mrs. Cecile White**, Director of Individual Development and Guidance, Horace Mann School, Teachers College, Columbia University, New York, New York.
- Foote, John M.**, Director, Reference and Service, State Department of Education, Baton Rouge, Louisiana.
- Foster, Richard R.**, Assistant Director, Research Division, National Education Association, Washington, D. C.
- Fowlkes, John G.**, Professor of Education, University of Wisconsin, Madison, Wisconsin.
- Franklin, E. E.**, Associate in Education, Johns Hopkins University, Baltimore, Maryland.
- Freeman, Frank N.**, Professor of Educational Psychology, Department of Education, University of Chicago, Chicago, Illinois.
- Fritz, Ralph A.**, Professor of Education, Kansas State Teachers College, Pittsburg, Kansas.
- Frutchey, Fred P.**, Research Associate, Bureau of Educational Research, Ohio State University, Columbus, Ohio.
- Fulk, Joseph Richard**, Professor of Public School Administration, College of Education, University of Florida, Gainesville, Florida.
- Gambrell, Bessie L.**, Associate Professor, Elementary Education, Yale University, New Haven, Connecticut.

Ganders, Harry S., Dean, School of Education, Syracuse University, Syracuse, New York.

Gans, Roma, Associate in Elementary Education, Teachers College, Columbia University, New York, New York.

Garrison, K. C., Professor of Psychology, University of North Carolina, Raleigh, North Carolina.

Garrison, S. C., Professor of Educational Psychology, and Dean of the Graduate School, George Peabody College for Teachers, Nashville, Tennessee.

Garver, F. M., Professor of Elementary Education, University of Pennsylvania, Philadelphia, Pennsylvania.

Gates, Arthur I., Professor of Education, Teachers College, Columbia University, New York, New York.

Gatto, Frank M., Assistant Director, Curriculum Study and Research, Pittsburgh Public Schools, Pittsburgh, Pennsylvania.

Gerberich, J. R., Assistant to Director, Education Division, Works Progress Administration, Washington, D. C.

Geyer, Denton L., Head of Department of Education and Psychology, Chicago Normal College, Chicago, Illinois.

Gifford, C. W., Chairman, Department of Psychology, Wright City Junior College, Chicago, Illinois.

Gillet, Harry O., Principal, Elementary School, University of Chicago, Chicago, Illinois.

Gilmore, Charles H., Director of Research, State Department of Education, Nashville, Tennessee.

Glenn, Earl R., Head of Science Department, New Jersey State Teachers College, Montclair, New Jersey.

Goldthorpe, J. Harold, Professor of Education, University of Rochester, Rochester, New York.

Good, Carter V., Professor of Education, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Goodrich, T. V., Director of Research, Public Schools, Lincoln, Nebraska.

Goodykoontz, Bess, Assistant Commissioner of Education, United States Office of Education, Washington, D. C.

Gordon, Hans C., Special Assistant to the Director of Educational Research, Philadelphia Public Schools, Philadelphia, Pennsylvania.

Grace, Alonzo G., Assistant Director, Extension, University of Rochester, Rochester, New York.

Gray, C. T., Professor of Educational Psychology, University of Texas, Austin, Texas.

Gray, Howard A., Research Associate, Erpi Pictures Consultants, Inc., New York, New York.

Gray, Robert Floyd, Director, Bureau of Research, Adult Education and Evening Schools, Board of Education, San Francisco, California.

Gray, William S., Professor of Education and Executive Secretary of the Committee on the Preparation of Teachers, Department of Education, University of Chicago, Chicago, Illinois.

Greenberg, Benjamin B., Assistant Superintendent of Schools, New York, New York.

Greene, H. A., Director, Bureau of Educational Research, Extension Division, University of Iowa, Iowa City, Iowa.

Greene, Katharine B., Assistant Professor of Psychology, School of Education, University of Michigan, Ann Arbor, Michigan.

Gregory, Marshall, Director, Division of Finance and Research, State Department of Public Instruction, Oklahoma City, Oklahoma.

Grossnickle, Foster E., Professor of Mathematics, State Teachers College, Jersey City, New Jersey.

Grover, Elbridge C., Superintendent of Schools, Euclid, Ohio.

Guiler, Walter S., Professor of Education and Director of Remedial Instruction, Miami University, Oxford, Ohio.

Haggerty, M. E., Dean, College of Education, University of Minnesota, Minneapolis, Minnesota.

Hall, Clifton W., Placement Adviser, Adelbert College, Western Reserve University, Cleveland, Ohio.

Hall, Sidney B., State Superintendent of Public Instruction, State Board of Education, Richmond, Virginia.

- Hanna, Paul R.**, Associate Professor, Stanford University, California.
- Hanson, Whittier L.**, Professor of Education, School of Education, Boston University, Boston, Massachusetts.
- Harap, Henry**, Associate Professor of Education, Western Reserve University, Cleveland, Ohio.
- Harrington, H. L.**, Supervising Director of Intermediate Schools, Board of Education, Detroit, Michigan.
- Harry, David P., Jr.**, Associate Professor of Education, Graduate School, Western Reserve University, Cleveland, Ohio.
- Hartmann, George W.**, Professor of Psychology, Pennsylvania State College, State College, Pennsylvania.
- Heaton, Kenneth L.**, Director, Division of Curriculum Research, State Department of Public Instruction, Lansing, Michigan.
- Heck, Arch O.**, Professor of Education, Ohio State University, Columbus, Ohio.
- Heilman, J. D.**, Director of Personnel Department, Colorado State Teachers College, Greeley, Colorado.
- Henmon, V. A. C.**, Department of Psychology, University of Wisconsin, Madison, Wisconsin.
- Henry, Mary Bess**, Counselor, Manual Arts High School, Los Angeles, California.
- Henry, Nelson B.**, Associate Professor of Education, School of Education, University of Chicago, Chicago, Illinois.
- Hertzberg, Oscar Edward**, Head, Department of Psychology, and Director of Research, State Teachers College, Buffalo, New York.
- Hertzler, Silas**, Professor of Education and Registrar, Goshen College, Goshen, Indiana.
- Hicks, J. Allan**, Professor of Education, New York State College for Teachers, Albany, New York.
- Hildreth, Gertrude**, Psychologist, Lincoln School of Teachers College, Columbia University, New York, New York.
- Hockett, John A.**, Department of Education, University of California, Berkeley, California.
- Hoke, K. J.**, Dean, College of Education, College of William and Mary, Williamsburg, Virginia.
- Hollingworth, Leta S.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Holy, T. C.**, Bureau of Educational Research, Ohio State University, Columbus, Ohio.
- Hopkins, L. Thomas**, Associate Professor of Education, Teachers College, Columbia University, New York, New York.
- Horan, Ellamay**, Professor of Education, De Paul University, Chicago, Illinois.
- Horn, Ernest**, Professor of Education, State University of Iowa, Iowa City, Iowa.
- Hubbard, Frank W.**, Associate Director, Research Division, National Education Association, Washington, D. C.
- Hughes, W. Hardin**, Director, Bureau of Administrative Research, Public Schools, Pasadena, California.
- Hurd, A. W.**, Associate Professor of Education, Northern Montana College, Havre, Montana.
- Irby, Nolen M.**, State Supervisor of Colored Schools, State Department of Education, Little Rock, Arkansas.
- Irwin, Manley E.**, Director, Department of Instruction, Public Schools, Detroit, Michigan.
- Jacobs, Clara M.**, Director of Educational Research, Centennial High School Building, Pueblo, Colorado.
- Jensen, Kai**, Assistant Professor of Education, University of Wisconsin, Madison, Wisconsin.
- Jersild, Arthur T.**, Associate Professor of Education, Teachers College, Columbia University, New York, New York.
- Jessen, Carl A.**, Senior Specialist in Secondary Education, United States Office of Education, Washington, D. C.
- Job, Leonard B.**, President, Ithaca College, Ithaca, New York.
- Johnson, George R.**, Director, Division of Tests and Measurements, Board of Education, St. Louis, Missouri.
- Johnson, J. T.**, Head, Department of Mathematics, Chicago Normal College, Chicago, Illinois.

Johnson, Palmer O., Associate Professor of Education, College of Education, University of Minnesota, Minneapolis, Minnesota.

Jones, Arthur J., Professor of Secondary Education, University of Pennsylvania, Philadelphia, Pennsylvania.

Jones, Harold E., Director, Institute of Child Welfare, University of California, Berkeley, California.

Jordan, A. M., Professor of Educational Psychology, University of North Carolina, Chapel Hill, North Carolina.

Jorgensen, A. N., President, Connecticut State College, Storrs, Connecticut.

Keeler, Louis Ward, Associate Professor of Educational Psychology, University of Michigan, Ann Arbor, Michigan.

Keener, E. E., Principal, Hay School, Chicago, Illinois.

Kelley, Truman L., Professor of Education, Graduate School of Education, Harvard University, Cambridge, Massachusetts.

Kelley, Victor H., Director, Teacher Training, Arizona State Teachers College, Flagstaff, Arizona.

Kelly, Fred J., Chief, Division of Higher Education, United States Office of Education, Washington, D. C.

Kemmerer, W. W., Director of Child Accounting and Curriculum, Independent School District, Houston, Texas.

Keys, Noel, Associate Professor of Education, University of California, Berkeley, California.

Kingsley, John H., Director, Division of Research, Board of Education, Albany, New York.

Kirby, T. J., Professor of Education, College of Education, State University of Iowa, Iowa City, Iowa.

Knight, F. B., Professor of Psychology and Education, University of Iowa, Iowa City, Iowa.

Knudsen, C. W., Professor of Secondary Education, George Peabody College for Teachers, Nashville, Tennessee.

Koch, Harlan C., Assistant Director, Bureau of Cooperation, University of Michigan, Ann Arbor, Michigan.

Koos, L. V., Professor of Secondary Education, University of Chicago, Chicago, Illinois.

Kramer, Grace A., Baltimore Public Schools, Baltimore, Maryland.

Kyte, George C., Professor of Education, University of California, Berkeley, California.

La Salle, Jessie, Assistant Superintendent in Charge of Educational Research, D. C. Public Schools, Washington, D. C.

Latham, O. R., President, Iowa State Teachers College, Cedar Falls, Iowa.

Lathrop, Frank W., Specialist in Agricultural Education, United States Office of Education, Washington, D. C.

Lee, J. Murray, Director of Research, Burbank City Schools, Burbank, California.

Lehman, Harvey C., Professor of Psychology, Ohio University, Athens, Ohio.

Leonard, J. Paul, Professor of Education, College of William and Mary, Williamsburg, Virginia.

Lide, Edwin S., Sullivan High School, Chicago, Illinois.

Lincoln, Edward A., Consulting Psychologist, Harvard University, Cambridge, Massachusetts.

Lindquist, E. F., Associate Professor of Education, State University of Iowa, Iowa City, Iowa.

Linn, Henry H., Business Manager, Board of Education, Muskegon, Michigan.

Loomis, Arthur K., Principal, University High School, University of Chicago, Chicago, Illinois.

Lovejoy, Philip, First Assistant Secretary, Rotary International, Chicago, Illinois.

MacLachy, Josephine, Bureau of Educational Research, Ohio State University, Columbus, Ohio.

Madsen, I. N., Director, Department of Tests and Measurements, State Normal School, Lewiston, Idaho.

Maller, Julius B., Research Associate, Teachers College, Columbia University, New York, New York.

Mallory, Clara, Professor of Education, Lamar Junior College, Beaumont, Texas.

Mann, Carleton H., Lecturer in Education, University of Southern California, Los Angeles, California.

- Manuel, H. T.**, Professor of Educational Psychology, University of Texas, Austin, Texas.
- Masters, Harry V.**, Superintendent, Ohio University Consolidated Training Schools, The Plains, Ohio.
- Mathews, C. O.**, Professor of Education, Ohio Wesleyan University, Delaware, Ohio.
- May, Mark A.**, Director, Institute of Human Relations, Yale University, New Haven, Connecticut.
- McCall, William A.**, Teachers College, Columbia University, New York, New York.
- McClure, Worth**, Superintendent of Schools, Seattle, Washington.
- McLaughlin, Katherine L.**, Associate Professor of Education, University of California at Los Angeles, Los Angeles, California.
- McLure, John R.**, Professor of School Administration, and Director of Summer School, University of Alabama, University, Alabama.
- Mead, A. R.**, Director of Educational Research, University of Florida, Gainesville, Florida.
- Meek, Lois Hayden**, Director, Child Development Institute, Teachers College, Columbia University, New York, New York.
- Melby, Ernest O.**, Dean, School of Education, Northwestern University, Evanston, Illinois.
- Melcher, George**, Superintendent of Schools, Kansas City, Missouri.
- Mendenhall, James E.**, Lincoln School of Teachers College, Columbia University, New York, New York.
- Meriam, Junius L.**, Professor of Education, University of California, Los Angeles, California.
- Merriman, Curtis**, Professor of Education, University of Wisconsin, Madison, Wisconsin.
- Miller, Chester F.**, Superintendent of Schools, Saginaw, Michigan.
- Miller, W. S.**, Professor of Educational Psychology, University of Minnesota, Minneapolis, Minnesota.
- Moehlman, A. B.**, School of Education, University of Michigan, Ann Arbor, Michigan.
- Monroe, W. S.**, Director, Bureau of Educational Research, University of Illinois, Urbana, Illinois.
- Moore, Clyde B.**, Professor in the Graduate School of Education, Cornell University, Ithaca, New York.
- Morgan, Walter E.**, Assistant Superintendent of Public Instruction, State Department of Education, Sacramento, California.
- Morphet, Edgar L.**, Director, Division of Research and Information, State Department of Education, Montgomery, Alabama.
- Morphett, Mabel Vogel**, Director of Research, Skokie School, Winnetka, Illinois.
- Morrison, J. Cayce**, Assistant Commissioner for Elementary Education, State Department of Education, Albany, New York.
- Mort, Paul R.**, Director of the Advanced School of Education, Teachers College, Columbia University, New York, New York.
- Morton, R. L.**, Professor of Mathematics, College of Education, Ohio University, Athens, Ohio.
- Mosher, Raymond M.**, Professor of Psychology, State College, San Jose, California.
- Myers, Anna G.**, Assistant Director of Research, Public Schools, Kansas City, Missouri.
- Myers, Charles Everett**, Supervisor, Division of Research, State Board of Education, Richmond, Virginia.
- Myers, Garry C.**, Head, Department of Parent Education, Cleveland College, Western Reserve University, Cleveland, Ohio.
- Nash, Harry B.**, Superintendent of Schools, West Allis, Wisconsin.
- Nelson, M. J.**, Dean of the Faculty, Iowa State Teachers College, Cedar Falls, Iowa.
- Nelson, Milton G.**, Dean, New York State College for Teachers, Albany, New York.
- Newland, T. Ernest**, Assistant Professor of Education, Bucknell University, Lewisburg, Pennsylvania.
- Nifenecker, Eugene A.**, Director, Bureau of Reference, Research, and Statistics, Board of Education, New York, New York.
- Noble, Stuart G.**, Professor of Education, Tulane University, New Orleans, Louisiana.
- Norton, John K.**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Norton, Mrs. John K.**, 464 Riverside Drive, Apt. 91, New York, New York.

Obrien, F. P., Director, Bureau of School Service and Research, University of Kansas, Lawrence, Kansas.
Odell, C. W., Associate Professor of Education, University of Illinois, Urbana, Illinois.
Ojemann, R. H., Assistant Professor, Iowa Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.
Olson, W. C., Director of Research in Child Development and Professor of Education, School of Education, University of Michigan, Ann Arbor, Michigan.
Oppenheimer, J. J., Dean of College of Liberal Arts, University of Louisville, Louisville, Kentucky.
O'Rear, F. B., Associate Professor of Education, Teachers College, Columbia University, New York, New York.
Orleans, Jacob S., Assistant Professor of Education, College of the City of New York, New York, New York.
O'Rourke, L. J., Director of Research in Personnel Administration, United States Civil Service Commission, Washington, D. C.
Osburn, W. J., State Teachers College, Buffalo, New York.
Otis, Arthur S., Psychological Editor, World Book Company, Yonkers, New York.
Otto, Henry J., Director of Education, W. K. Kellogg Foundation, Battle Creek, Michigan.
Paul, Joseph B., Director, Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa.
Peik, W. E., Professor of Education, University of Minnesota, Minneapolis, Minnesota.
Perry, Winona M., Professor of Educational Psychology and Measurements, University of Nebraska, Lincoln, Nebraska.
Peters, Charles C., Director of Educational Research, Pennsylvania State College, State College, Pennsylvania.
Peterson, Elmer T., Associate Professor of Education, College of Education, University of Iowa, Iowa City, Iowa.
Phillips, Albert J., Deputy Executive Secretary, Michigan Education Association, Lansing, Michigan.
Potter, Mary A., Supervisor of Mathematics, Washington Park High School, Racine, Wisconsin.
Potthoff, Edward F., Assistant Professor of Education, University of Illinois, Urbana, Illinois.
Powers, S. R., Professor of Natural Sciences, Teachers College, Columbia University, New York, New York.
Prall, Charles E., Dean, School of Education, University of Pittsburgh, Pittsburgh, Pennsylvania.
Prescott, D. A., Professor of Education, Rutgers University, New Brunswick, New Jersey.
Pressey, S. L., Professor of Educational Psychology, College of Education, Ohio State University, Columbus, Ohio.
Price, Malcolm P., Chairman, Personnel Committee, Detroit Public Schools, Detroit, Michigan.
Proffitt, Maris M., Educational Consultant and Specialist in Guidance and Industrial Education, United States Office of Education, Washington, D. C.
Rankin, Paul T., Director of Curriculum and Research, Board of Education, Detroit, Michigan.
Reavis, W. C., Professor of Education, University of Chicago, Chicago, Illinois.
Reeder, Ward G., Professor of School Administration, Ohio State University, Columbus, Ohio.
Reeves, Floyd W., University of Chicago, Chicago, Illinois.
Remmers, H. H., Professor of Education and Psychology, Purdue University, Lafayette, Indiana.
Reusser, Walter C., Professor of Education, University of Wyoming, Laramie, Wyoming.
Richey, Herman G., Assistant Professor of Education, University of Chicago, Chicago, Illinois.
Rinsland, H. D., Associate Professor of Education, Bureau of Educational Research, University of Oklahoma, Norman, Oklahoma.
Rogers, Don C., Director, Bureau of Research and Building Survey, Board of Education, Chicago, Illinois.

- Rosenlof, George W.**, Professor of Secondary Education, University of Nebraska, Lincoln, Nebraska.
- Rowland, W. T., Jr.**, Assistant Superintendent in Charge of Secondary Education, Louisville Public Schools, Louisville, Kentucky.
- Ruch, G. M.**, Scott, Foresman and Company, Chicago, Illinois.
- Rugg, Earle U.**, Head, Division of Education, Colorado State College of Education, Greeley, Colorado.
- Rugg, Harold**, Professor of Education, Teachers College, Columbia University, New York, New York.
- Rulon, Phillip J.**, Assistant Professor, Graduate School of Education, Harvard University, Cambridge, Massachusetts.
- Russell, John Dale**, Associate Professor of Education, University of Chicago, Chicago, Illinois.
- Sackett, Everett B.**, Associate in Research, Lincoln School of Teachers College, Columbia University, New York, New York.
- Sanchez, George I.**, Director, Division of Information and Statistics, State Department of Education, Santa Fe, New Mexico.
- Sangren, Paul V.**, Dean of Administration, Western State Teachers College, Kalamazoo, Michigan.
- Sawyer, Guy E.**, Chadds Ford, Pennsylvania.
- Seates, Douglas E.**, Director of School Research, Cincinnati Public Schools, Cincinnati, Ohio.
- Schorling, Raleigh**, Professor of Education and Director of Instruction, University High School, University of Michigan, Ann Arbor, Michigan.
- Schrammel, H. E.**, Director, Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.
- Sears, Jesse B.**, Professor of Education, Stanford University, California.
- Segel, David**, Specialist, Tests and Measurements, United States Office of Education, Washington, D. C.
- Senour, A. C.**, Assistant Superintendent, Public Schools, East Chicago, Indiana.
- Shea, James T.**, Director, Curriculum and Research, Board of Education, San Antonio, Texas.
- Simpson, Alfred D.**, Assistant Commissioner of Education, for Finance, State Education Department, Albany, New York.
- Simpson, B. R.**, Professor of Educational Psychology, Western Reserve University, Cleveland, Ohio.
- Sims, Verner M.**, Associate Professor of Psychology, College of Education, University of Alabama, University, Alabama.
- Singleton, Gordon G.**, Dean of Education and Director of Summer Quarter, Mercer University, Macon, Georgia.
- Smith, Dora V.**, Associate Professor in Education, College of Education, University of Minnesota, Minneapolis, Minnesota.
- Smith, H. L.**, Dean, School of Education, Indiana University, Bloomington, Indiana.
- Smith, Harry P.**, Professor of Education, Syracuse University, Syracuse, New York.
- Snyder, Agnes**, Associate in New College, Teachers College, Columbia University, New York, New York.
- Soper, Wayne W.**, Research Associate, State Education Department, Albany, New York.
- Spaulding, Francis T.**, Associate Professor of Education, Harvard University, Cambridge, Massachusetts.
- Spencer, Peter L.**, Professor of Education, Claremont Colleges, Claremont, California.
- Starbuck, Edwin D.**, Director of the Institute of Character Research, University of Southern California, Los Angeles, California.
- Stenquist, John L.**, Director, Bureau of Educational Research, Baltimore Public Schools, Baltimore, Maryland.
- Stern, Bessie C.**, Statistician, State Department of Education, Baltimore, Maryland.
- Stevenson, Fred Gray**, 1417 Lake Shore Drive, Muskegon, Michigan.
- Stoddard, George D.**, Director, Iowa Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.
- Stoke, Stuart M.**, Head of Education Department, Mount Holyoke College, South Hadley, Massachusetts.
- Stokes, C. Newton**, Department of Secondary Education, Temple University, Philadelphia, Pennsylvania.

Strachan, Lexie, Psychologist, Public Schools, Kansas City, Missouri.
Strang, Ruth M., Assistant Professor of Education, Teachers College, Columbia University, New York, New York.
Stratemeyer, Florence B., Associate Professor of Education, Teachers College, Columbia University, New York, New York.
Strayer, George D., Director, Division of Field Studies, Institute of Educational Research, Teachers College, Columbia University, New York, New York.
Streitz, Ruth, Professor of Education, University of Cincinnati, Cincinnati, Ohio.
Sumstine, David R., Director, Department of Curriculum Study and Research, Public Schools, Pittsburgh, Pennsylvania.
Sutton, D. H., Director, Division of School Finance, State Department of Education, Columbus, Ohio.
Swift, Fletcher Harper, Professor of Education, University of California, Berkeley, California.
Symonds, Percival M., Professor of Education, Teachers College, Columbia University, New York, New York.
Terry, Paul W., Professor of Educational Psychology, University of Alabama, University, Alabama.
Theisen, W. W., Assistant Superintendent of Schools, Milwaukee, Wisconsin.
Thurber, Clarence Howe, President, University of Redlands, Redlands, California.
Tidwell, Robert E., Director of Extension, University of Alabama, University, Alabama.
Tiegs, Ernest W., Dean, University College, University of Southern California, Los Angeles, California.
Tilton, J. Warren, Associate Professor of Educational Psychology, Department of Education, Yale University, New Haven, Connecticut.
Tink, Edmund L., Superintendent of Schools, Kearny, New Jersey.
Toops, Herbert A., Professor of Psychology, Department of Psychology, Ohio State University, Columbus, Ohio.
Torgerson, T. L., Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.
Tormey, T. J., President, Arizona State Teachers College, Flagstaff, Arizona.
Touton, Frank C., Vice President and Professor of Educational Research, University of Southern California, Los Angeles, California.
Townsend, M. Ernest, President, State Normal School, Newark, New Jersey.
Trabue, M. R., Director, Occupational Research Program, United States Employment Service, Department of Labor, Washington, D. C.
Trow, William Clark, Professor of Educational Psychology, University of Michigan, Ann Arbor, Michigan.
Turney, Austin Henry, University of Kansas, Lawrence, Kansas.
Tyler, Ralph W., Professor of Education, Bureau of Educational Research, Ohio State University, Columbus, Ohio.
Tyler, Tracy Ferris, Secretary and Research Director, National Committee on Education by Radio, Washington, D. C.
Uhl, Willis L., Professor of Education and Dean, School of Education, University of Washington, Seattle, Washington.
Umstadtd, J. G., Assistant Professor of Education, University of Minnesota, Minneapolis, Minnesota.
Updegraff, Harlan, Educational Consultant, Stratt Haven Inn, Swarthmore, Pennsylvania.
Upshall, Charles Cecil, Director, Bureau of Research, State Normal School, Bellingham, Washington.
Van Wagenen, M. J., Assistant Professor of Educational Psychology, University of Minnesota, Minneapolis, Minnesota.
Vreeland, Wendell, Associate Professor of Education, Wayne University, Detroit, Michigan.
Walker, Helen M., Assistant Professor of Education, Teachers College, Columbia University, New York, New York.
Waples, Douglas, Professor of Educational Method, Graduate Library School, University of Chicago, Chicago, Illinois.
Washburne, Carleton W., Superintendent of Schools, Winnetka, Illinois.
Washburne, John N., Associate Professor of Educational Psychology, Syracuse University, Syracuse, New York.

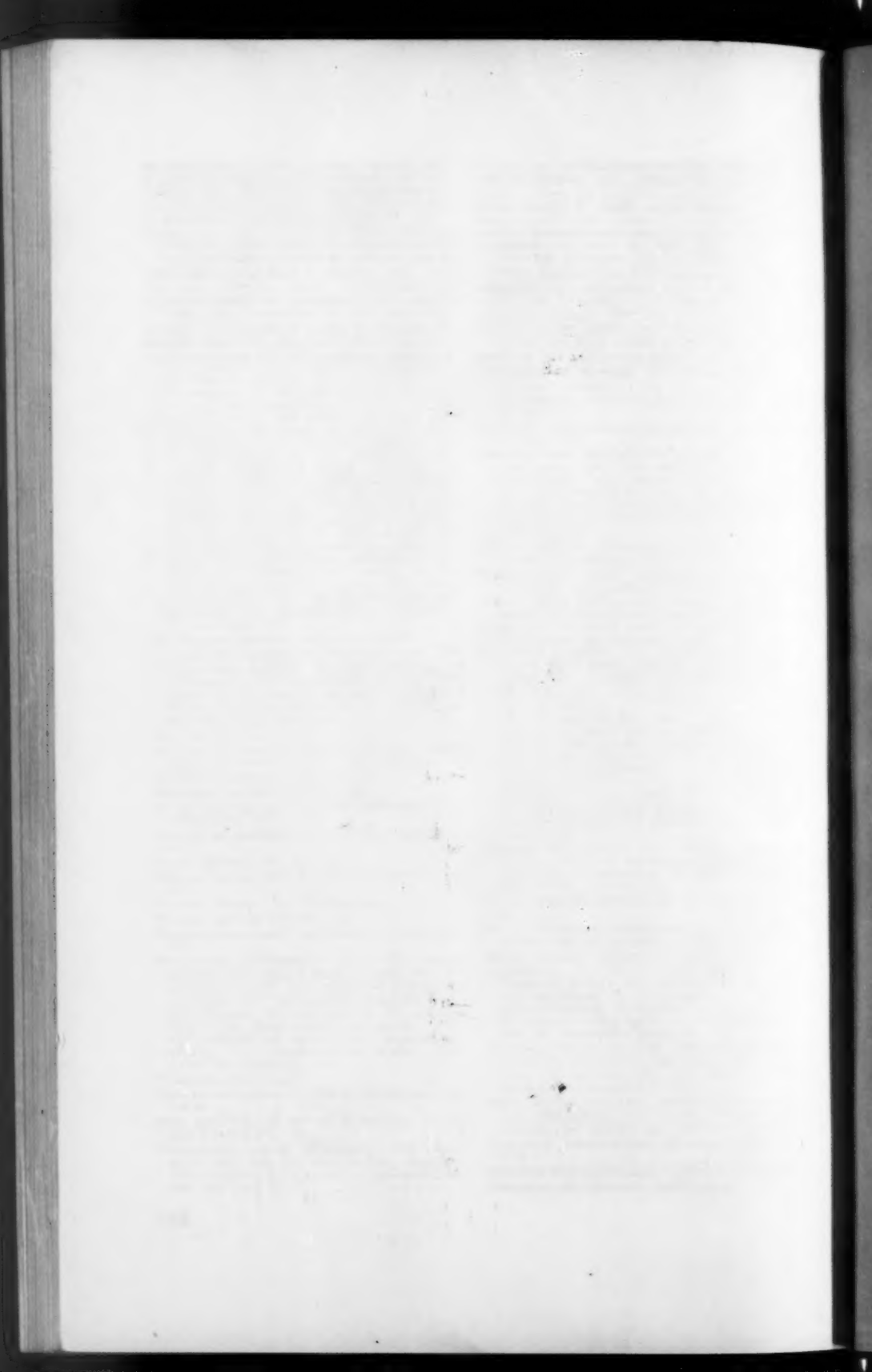
- Waterman, Ivan R.**, Chief, Division of Textbooks and Publications, California State Department of Education, Sacramento, California.
- Watkins, Ralph K.**, Professor of Education, School of Education, University of Missouri, Columbia, Missouri.
- Watson, Goodwin**, Associate Professor of Education, Teachers College, Columbia University, New York, New York.
- Welles, J. B.**, Principal, State Normal School, Geneseo, New York.
- West, Paul V.**, Professor of Education, New York University, New York, New York.
- Wheat, Harry G.**, Professor of Education, West Virginia University, Morgantown, West Virginia.
- Wilber, Flora**, Director, Research and Measurement, Fort Wayne Public Schools, Fort Wayne, Indiana.
- Williams, J. Harold**, Professor of Education, University of California at Los Angeles, Los Angeles, California.
- Willing, M. H.**, Professor of Education, University of Wisconsin, Madison, Wisconsin.
- Wilson, Guy M.**, Professor of Education, School of Education, Boston University, Boston, Massachusetts.
- Wilson, W. K.**, Assistant, School Buildings and Grounds Division, State Education Department, Albany, New York.
- Witham, Ernest C.**, Associate Professor of Education, Rutgers University, New Brunswick, New Jersey.
- Witty, Paul A.**, Professor of Education, School of Education, Northwestern University, Evanston, Illinois.
- Wood, Ben D.**, Associate Professor of Collegiate Research, Columbia University, New York, New York.
- Wood, E. R.**, Associate Professor of Psychology, School of Education, New York University, New York, New York.
- Woods, Elizabeth L.**, Supervisor, Educational Research and Guidance Section, Chamber of Commerce, Los Angeles, California.
- Woods, Roy C.**, Professor of Education, Marshall College, Huntington, West Virginia.
- Woody, Clifford**, Director, Bureau of Educational Reference and Research, University of Michigan, Ann Arbor, Michigan.
- Worcester, D. A.**, Head, Department of Educational Psychology and Measurements, University of Nebraska, Lincoln, Nebraska.
- Wrenn, C. Gilbert**, Assistant Registrar and Director of Vocational Guidance, Stanford University, California.
- Wright, Wendell W.**, Professor of Education, Indiana University, Bloomington, Indiana.
- Wrightstone, J. Wayne**, Research Associate, Teachers College, Columbia University, New York, New York.
- Yates, Mrs. Dorothy H.**, Associate Professor of Psychology, San Jose State College, San Jose, California.
- Yeager, William A.**, Professor of Administration, University of Pittsburgh, Pittsburgh, Pennsylvania.
- Young, William E.**, Supervisor, Intermediate Grades, Hibbing, Minnesota.
- Zirbes, Laura**, Professor of Education, Ohio State University, Columbus, Ohio.

INDEX TO VOLUME V

- Accounting, public school, administration, 125; bibliography, 171; internal or non-public, 126; procedure and principles, 125; raw cost appraisal, 124; state systems of, 124**
- Achievement, scholastic, personal data used in predicting, 223; validities of tests, 223; see also Educational tests**
- American Educational Research Association, list of members, 521; officers for 1935-36 (inside front cover, all issues); qualifications for membership (inside front cover, all issues); relationship to the National Education Association (inside front cover, all issues)**
- Aptitude, bibliography, 305; common fallacies concerning, 215; criteria of tests of, 217; evaluations of test programs, 218; inadequacy in testing, 216; measures of, 215; scholastic tests in specific fields, 224; statewide cooperative testing programs, 219; tests for, in foreign countries, 452; tests of scholastic, 221**
- Arithmetic, bibliography, 93; diagnosis and remedial instruction, 21; early number experience, 14; level of attainment in college, 23; mental age and achievement in, 14; methods of teaching, 16; permanence of improvement in, 22; problem solving, 24; selection and arrangement of subjectmatter, 25; sex differences in, 26; transfer of training, 20; vocabulary studies, 27**
- Arts, fine, bibliography, 104; influence of training on ability in, 46; methods of teaching and recent trends, 45**
- Bibliographies, achievement tests in colleges and universities, 517; activities in the nursery school, kindergarten, and elementary grades, 89; applications of intelligence testing, 296; arithmetic, 93; capital outlay, indebtedness, and debt service, 177; character education, 97; educational tests and measurements in China, England, France, and Germany, 502; English, 99; equipment, apparatus, and supplies, 421; finance and business administration in institutions of higher education, 177; financial economy and school business administration, 181; fine arts, 104; health and physical education, 106; general survey of character and personality measurement, 315; heating, ventilation, and sanitation in school buildings, 413; intelligence and its measurement, 291; measures of aptitude, 305; measures of character and personality through conduct and information, 325; mental hygiene and emotional adjustment, 315; needed research in the field of school buildings and equipment, 439; objective achievement test construction, 513; operation and care of the school plant, 427; plant development for higher education, including junior colleges, 428; present tendencies in the uses of educational measurements, 510; public education costs, 180; public school accounting, 171; public school budget, 171; public school plant insurance, 426; reading, 111; recent developments in the written essay examination, 516; recent trends in school-plant planning, 437; research and survey technics, 176; revenues and taxation, 172; school playgrounds: their surfacing, administration, use and care, 425; science, 114; social attitudes, 320; social studies, 115; spelling, 118; state and federal aid, 183; technics for determining housing requirements in elementary, junior, and senior high schools, 412; test construction and statistical interpretation, 309**
- Budgeting, public school, administration and organization, 128; appraisal, 130; bibliography, 171; effect of depression on, 130; for school plants, 129; in miscellaneous items, 156; principles, 127**
- Buildings, school. See School plant**
- Business administration. See Finance and business administration**
- Capital outlay, indebtedness, and debt service, bibliography, 177; studies of, 141**
- Character. See Personality**
- Character education, bibliography, 97; classroom methods, 33; field and methods of attack, 31; studies of non-classroom methods, 35**
- Character tests, used in foreign countries, 452**
- Clerical aptitude, tests for, 225**
- Costs of education, bibliography, 180; public school, 148**
- Debt service. See Capital outlay, indebtedness, and debt service**
- Delinquency, measurement of tendencies, 278**
- Dentistry, aptitude tests for, 224**
- Economy, in relation to the administrative unit, 151; in school business administration, 151**
- Education, bibliography on costs, 180; economic value of, 169**
- Educational tests. See Tests, educational**
- Elementary school, activity as method in, 4; development of social behavior, 6; learning and language, 11; subjectmatter, 12; use of art materials, books, and music, 6**
- Emotional adjustment, bibliography, 315; physiological and laboratory tests of, 253; testing instruments for self-report, 245; traits tested by behavior, 251; treatment of, correlated with tests, 254**
- Engineering, aptitude tests for, 224**
- English language, bibliography, 99; content and placement of the curriculum, 38; evaluation of method, 41; measurement of achievement, 43; remedial instruction, 43; research technics, 37**
- Equalization of educational opportunity, research methods for studying, 140**
- Equipment, apparatus, and supplies, administration of, 153; bibliography, 421; for operating school plant, 380; for playgrounds, 366; trends in, 362**
- Examinations and marks, in foreign countries, 451; recent development in written essay, 484; see also Tests, educational**
- Federal aid for education, bibliography, 183; studies of, 158, 164**
- Feeblemindedness, studies of, 193, 205**
- Finance and business administration, bibliography, 177, 181; in institutions of higher education, 143; outlook for future research, 165; studies of, 124-84**
- Fine arts. See Arts, fine**
- Health and physical education, bibliography, 106; diagnosis and individual needs, 48; learning process involved in, 49; methods of teaching, 52; motor ability and athletic skills, 50; physical defects, 50; relation to motor ability and other factors, 49; tests and measurements, 51; see also Playgrounds, school**

- Heating, ventilation, and sanitation in school buildings**, bibliography, 413; lighting, 357; recent studies of, 348; review of theories concerning, 344; sanitary equipment, 354; *see also* Equipment, apparatus, and supplies
- Higher education**, bibliography on school plant development for, 428; economy in finance and business administration, 143; school plant development for, 388
- Insurance of school plant**, bibliography, 426; city plans, 371; comparison of premiums and losses, 372; present status of legislation regulating, 370; reducing costs, 374; state plans, 371
- Intelligence**, applications of testing, 199; as related to environment, 210; as related to racial characteristics, 212; bibliography on applications of testing, 296; bibliography on measurement of, 291, 305; clinical interpretations of, 189; constancy of I. Q., 191; correlations of tests of, with college scholarship, 222; foreign tests, 197; group test equivalent scores, 195; growth in, 191; individual differences in, 207; measurement of, 187; nature of, 187; new and revised group tests of, 196; new and revised individual tests of, 196; preschool studies of, 192; qualitative phases of testing, 190; relation to other traits, 209; speed as a factor in, 189; studies of gifted and superior children, 192, 206; studies of the feeble-minded, 193, 205; surveys and interpretations, 200; test comparability, 195; testing delinquents, prisoners, and criminals, 204; testing of, for scholastic purposes, 201; testing of, in vocational guidance, 207; test standardization and variability, 194; variability of, at adult levels and senescence, 193
- Intelligence quotient**, constancy of, 191
- Janitor, school**, status of, 378; training of, 379
- Kindergarten**, activities as special methods, 4; bibliography, 89; development of social behavior, 5; learning and language, 8; play activities, 7; subjectmatter, 10, 11; use of art materials, books, and music, 6
- Law**, aptitude tests for, 224
- Lighting of school building**. *See* Heating, ventilation, and sanitation in school buildings
- Mechanical aptitude**, tests for, 225
- Medicine**, aptitude tests for, 224
- Mental hygiene**, and emotional adjustment, 245; bibliography, 315
- Methods of teaching**, at elementary-school level, 4-120
- Music**, aptitude tests for, 225
- National Survey of School Finance**, study cited, 159
- Nursery school**. *See* Kindergarten
- Nursing**, aptitude tests for, 225
- Organization, school**, as related to instruction, 152
- Personality**, bibliography, 315, 325; factor analysis, 277; general survey of tests, 242; inventories of, 277; measurement of delinquency tendencies, 278; measurement of specific traits, 281; measures of, thru conduct and information, 273; measurement by ratings, 274; of emotional adjustment, 245; physiological studies, 276; psychogalvanic studies, 276; scales for measuring, 279
- Phonetics**, value of, 57
- Physical education**. *See* Health and physical education
- Play activities**, as method of teaching, 7; *see also* Playgrounds, school
- Playgrounds, school**, bibliography, 425; equipment, 366; size and location, 366; standards, 367; surfaces, 367; surfacing, administration, use, and care, 364
- Psychology of teaching**, at elementary-school level, 4-120
- Rating scales**, for measuring character and personality, 274
- Reading**, ability in, as related to intelligence, 63; appreciation of literature, 62; beginning reading methods compared, 56; bibliography, 111; causes of deficiencies, 64; children's preferences, 68; classifying and adapting instruction to pupils, 59; improving in content field, 61; individual and group differences, 58; interests, 67; methods of attacking words in primary reading, 57; methods of improving silent, compared, 60; readiness, 55; relation between rate and comprehension, 62; remedial procedures, 65; size of type and readability, 56; successive emphases in method, 54; systematic versus incidental training in, 55; value of phonetics, 57; value of various types of writing in beginning, 56
- Research organizations, educational**, in foreign countries, 444
- Revenues and taxation**, bibliography, 172; income taxes, 137; local taxing power versus centralization, 168; productiveness of education, 169; property tax, 136; tax delinquency and local school revenue, 132; tax relief and equalization, 133; trends in provisions of, 133; sales taxes, 137
- Salary schedules**, future research, 165; research methods for studying, 139
- Sanitation in school buildings**. *See* Heating, ventilation, and sanitation in school buildings
- School plant**, bibliography, general, 412; bibliography on operation and care of, 427; bibliography on recent trends in planning, 437; bibliography on suggested research on, 439; building standards, 339; capital outlay expenditures, 403; development of, for higher education, 388; flexibility and adaptation, 396; insurance, 370; management of, 153; operation and care of, 378; operating costs, 381; planning techniques, 340; recent trends in planning, 393; relation of city planning to planning for, 394; reviews of building plans, 339; school janitor, 378; special rooms, 337; standardization of requirements for, 400; suggested research, 406; technica for determining housing requirements of schools, 337; type of construction and materials as related to original cost, maintenance, and operation, 383; utilization of, 340
- Science**, bibliography, 114; curriculum studies in, 72; effects of learning on behavior, 73; psychology of instruction, 72; trends in teaching, 70, 71
- Social attitudes**, bibliography, 320; measurement of, 259
- Social behavior**, development in nursery and kindergarten, 5; psychological tests of, 259
- Social studies**, bibliography, 115; class size as related to efficiency, 76; curriculum, 77; influence of school's social setting, 76; materials and equipment, 77; measurements, 82; problems in learning, 78; problems in teaching, 80; reference to bibliographies on, 76; study methods, 79; vocabulary studies, 81
- Spelling**, bibliography, 118; deficiency and diagnosis, 87; measures of achievement, 87; references to bibliographies, 83; selection and gradation of vocabulary, 83
- State aid for education**, administrative aspects, 162; bibliography, 183; history, theories, evaluations, and technics, 163; studies of, 158
- Statistical interpretation of tests**, studies of, 229
- Supplies and equipment**, administration of, 155
- Taxation**. *See* Revenues and taxation

- Teaching**, aptitude tests for, 225
- Technics**, bibliography, 176; research and survey, 139
- Test administration**, effects of varying conditions, 227
- Test construction and interpretation**, bibliography, 309, 483; of objective achievement, 469; studies of, 229; trends in, 458
- Testing programs**, state and national, 456
- Tests, educational**, bibliography, 502; in colleges and universities, 491; influence of, 482; in foreign countries, 443, 445; present tendencies, 455; scoring of, 481; student reaction to, 488
- Tests, physical**, in foreign countries, 452
- Tests, psychological**, construction and statistical interpretation, 229; diagnostic and prognostic, in foreign countries, 452; measurement of mental functions, 487; of aptitude, 215; of character and personality, 242, 273; of intelligence, 187, 199; of social attitudes, 259; review of research relating to, 187, use of, in foreign countries, 453
- Transportation of pupils**, studies of, 156
- Type**, size of as related to readability, 56
- Unit costs**, methods of studying, 140; of public schools, 149
- Validity and reliability**, administrative factors affecting, 471; of achievement tests, 469
- Ventilation of school buildings**. *See* Heating, ventilation, and sanitation in school buildings
- Vocational guidance**, use of intelligence tests in, 207



REVIEW OF EDUCATIONAL RESEARCH

EACH issue of the REVIEW summarizes the research studies and scientific investigations bearing on a specific topic and includes extensive bibliographies. The content of the REVIEW is such that it is exceedingly useful to college professors and all students of education.

SUBSCRIPTIONS

Because of the fact that membership in the American Educational Research Association is by election only, it was found necessary to create a special subscription to take care of the popular demand for this publication.

Non-members of the American Educational Research Association and institutions may subscribe for the REVIEW at the rate of \$4.00 per year, or may purchase single copies at \$1.00 each.

RELATED ISSUES

Issues related to this REVIEW include: *Tests of Personality and Character*, June, 1932; *Tests of Intelligence and Aptitude*, October, 1932; *Methods and Technique of Educational Research*, February, 1934; *Psychology and Methods in the High School and College*, December, 1934; *Special Methods and Psychology in the Elementary School*, February, 1934; and *Psychological Tests*, June, 1935.

Topics for Forthcoming Issues of the Review, the Approximate Date Scheduled, and List of Committee Members

Mental and Physical Development of Children. February, 1936.

George D. Stoddard, Chairman
Fowler D. Brooks
Paycha Cattell
Harold E. Jones
Lois Hayden Meek

Student Accounting, Personnel, and Guidance. April, 1936.

Arch O. Heck, Chairman
Fred C. Ayer
A. J. Brumbaugh
H. H. Bixler
Ruth Strang

Psychology of Learning, General Methods, and Supervision. June, 1936.

G. T. Buswell, Chairman
W. A. Brownell
Ernest O. Melby
Prudence Cutright
Manley Irwin

History of Education and Comparative Education. October, 1936.

Newton Edwards, Chairman
Edward H. Reiser
I. L. Kandel
Stuart G. Noble

Mental Hygiene. December, 1936.

Harry J. Baker, Chairman
Percival M. Symonds
Willard C. Olson
J. Harold Williams
S. L. Prosser

Members of the American Educational Research Association are urged to write to the Chairman of the particular Review in which they may be primarily interested, sending reprints, other publications, and suggestions.

